UNDERSTANDING SOCIETY

Social Sciences Foundation Course Team

The Open University

James Barber
John Blunden
F. S. Brooman
Hedy Brown
Michael Drake
Ruth Finnegan
Robert Green
Alan Hancock
A. T. A. Learmonth
D. J. Murray
Jeremy Tunstall

Research Assistant: Joanna L. Taussig

UNDERSTANDING SOCIETY

Readings in the Social Sciences

Edited by the

Social Sciences Foundation Course Team

The Open University

Published by

Macmillan *for*

The Open University Press

Selection and editorial matter © The Open University 1970

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

First published 1970 by
MACMILLAN AND CO LTD
London and Basingstoke
Associated companies in New York Toronto
Dublin Melbourne Johannesburg and Madras

SBN 333 12227 5 (hard cover)

The Papermac edition of this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent, in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

Printed in Great Britain by
RICHARD CLAY (THE CHAUCER PRESS), LTD
Bungay, Suffolk

CONTENTS

List of Figures vii

List of Sources xi

Acknowledgements xv

Introduction xix

PART 1 **Why People Live in Societies**

1 The Fundamentals of Human Nature 3

2 Men and Government 23

3 Economic Co-operation 41

4 Societies and Environments 50

5 Man as a Social Animal 80

PART 2 **How People Live in Societies**

6 Socialisation 109

7 Personality Development 138

8 Attitudes and Prejudice 154

9 The Family and its Functions 175

10 Economic Wants 203

11 Production 208

12 Markets and Prices 215

13 The Sociology of Economic Behaviour 229

14 The Workings of the Economy 246

15 Money 256

16 Distribution of Incomes 264

17 Social Stratification 271

18 The Psychology of Social Class 276

19 Rural Land Use 289

20 Location of Industry 316

21 Zoning in Cities 340

22 The Size and Spacing of Settlements 362

23 Approaches to Political Geography 394

24 The Formal Structure of Government 424

25 Informal Political Institutions 440

26 Government and Politics without the State 472

27 Political Culture and Political Systems 491

28 Politics in Groups 525

29 Stability and Function in Society 559

30 Change in Society 567

31 Stability and Change in Groups 578

PART 3 **What Kinds of Problems People Face in Societies**

The Population Explosion: An Interdisciplinary Approach

32 The Demographer and his World 611

33 The Population Geography of the Indian Subcontinent during an Era of Change 620

34 Population and Economic Growth 643

35 Social and Political Implications of Population Increase 655

36 Diffusion and Acceptance of Change 674

Notes on Editors 694
Index of Authors 697
Index of Extract Titles 699
General Index 701

LIST OF FIGURES

4.1 Part of Chimbu Division, Eastern Highlands, New Guinea 51

4.2 The Mintima area, 1958–63 57

4.3 Group territories within a Chimbu tribe – Naregu 60

4.4 Some cultural boundaries; some ceremonial ground regions in part of central Chimbu 63

4.5 Some modern regions 64

4.6 (*a*) The central highlands of New Guinea and the Chimbu area 67

(*b*) Territorial divisions within the study area 67

(*c*) Altitude and soil-moisture conditions 67

4.7 A simplified model of the innovation path and selected associated changes 68

4.8 Land use in 1958 and 1962 70

4.9 Land use in 1965 and 1967 71

4.10 A Weberian model of coffee processing 72

4.11 Movement of coffee cherry from garden to processing point 74

4.11 (*a*) Movement of cherry to Iwagl central pulpery, 20 Jan–16 Apr 1965 74

(*b*) Approximate movement of cherry to processing points in May 1967 74

4.12 Some elements in the system of internal circulation and centrality 77

4.12 (*a*) Decomposition of the walking network by shortest path to the most connected men's house site, treating all tracks joining men's house gardens and women's houses for simplicity; distribution of the most connected sites 77

(*b*) A diagram of Mintima ceremonial ground in 1956, from information obtained in 1959 and analogy with grounds elsewhere in use in 1959 77

(*c*) Decomposition of the walking network by shortest path to Iwagl, following the most used tracks from men's house sites; distribution of some consumer-oriented functions; movement of central points of distributions 77

(*d*) Iwagl in 1967 77

5.1 The Nortons, spring and summer 1937 84

9.1 Plan of Woɖaaɓe homestead 187

14.1 Graph 247

14.2 Graph 248

14.3A Graph 249

14.3B Graph 250

14.3C Graph 250

14.4A Graph 250

14.4B Graph 251

14.5 Graph 251

19.1 The location of agricultural production after von Thünen: (*a*) The *isolierte Staat*; (*b*) modified conditions 290

19.2 Graph 291

19.3 Graph 292

19.4 United States freight-rate territories 294

19.5 Profile of railroad Class I freight rates, to/from Milwaukee, 1953 295

19.6 Wisconsin freight-rate structure, 1953: railroad first-class rates to/from Milwaukee, carload rates in cents per 100 lb. 296

19.7 Wisconsin freight-rate structure, 1953: coal rates to/from Milwaukee, carload rates in cents 296

19.8 Percentage of total farms classified as cash-grain, by counties 298

19.9 Percentage of total farms classified as cash-grain, by state economic areas 298

19.10 Percentage of total farm area in cash-grain farms by state economic areas 299
19.11 Percentage of flat land per county 299
19.12 Percentage of flat land by state economic areas 300
19.13 Error of the estimate of percentage of total farm area in cash-grain farms 301
19.14 The sample area, Middle Sweden, and its counties 303
19.15 Actual labour productivity: average per farm 304
19.16 Potential labour productivity: average per farm 304
19.17 Mean Productivity Index: ratio of actual to potential labour productivity 305
19.18 Productivity Index regions 305
19.19 The yield variability of feed grains 309
19.20 The yield variability of winter bread grains 309
19.21 The variability of crop yields 310
19.22 The variability of crop yields: deducted value per hectare 310
19.23 The von Thünen zones under several complicating conditions 313
20.1 Weberian weight-loss analysis of market or destination orientation (*A*, *B*) and material or origin orientation (*C*, *D*) 318
20.2 Distribution of sugar beet factories in relation to arable land 319
20.3 Milk manufacturers, 1934–5 320
20.4 Coke ovens, 1952 325
20.5 Open-hearth steel furnaces, 1953 327
20.6 Blast furnaces and open-hearth steel works, 1953 328
20.7 Location of plants constructed between 1945 and 1963 330
20.8 The Doboj project: the plan for industry in 1947 331
20.9 New plants constructed after the Cominform break, and the changed pattern of material movement 331
20.10 The distribution of industrial employment in the Republic of Ireland established with foreign participation between January 1959 and December 1966 333
20.11 Products in transit 337
21.1 General pattern of urban land values 341
21.2 The transformed land value maps, Chicago, 1910–40 343
21.3 The transformed land value maps, Chicago, 1950–60 344
21.4 Generalisations of internal structure of cities 349
21.5 Urban areas 350
21.6 Hypothetical rent–distance relationships within a city 351
21.7 The setting of Calgary and the relationship of the limits of the built-up area to terrain and transportation facilities 352
21.8 Sectors in the analysis 353
21.9 Hoyt's sector theory applied to Calgary's land use pattern, 1914, when the city population was approximately 50,000 354
21.10 Hoyt's sector theory applied to Calgary, 1948, when the metropolitan population was approximately 105,000 355
21.11 Hoyt's sector theory applied to Calgary, 1961, when the metropolitan population was approximately 270,000 356
21.12 Growth–distance relationships in urban conversion of land in southern Cambridgeshire 358
21.13 The distribution of commercial development in the Calgary metropolitan area, January 1961 359
22.1–4 Colonisation of a region of bankers 364
22.5 Efficiency of alternative types of regular polygons in relation to distance from centres and perimeter lengths 365

22.6 The system of central places after (*top*) the marketing principle, (*centre*) the administrative principle, (*bottom*) the transportation principle 367

22.7 Hamlets spaced two miles apart on an equilateral triangular grid 368

22.8 Pattern of hamlets and villages service area governed by a $k = 3$ hierarchy 368

22.9 Hamlets, villages and towns in a 'fixed-k' hierarchy where $k = 3$ (Christaller's 'marketing' principle) 369

22.10 Pattern of service areas of villages in a $k = 4$ hierarchy 369

22.11 Pattern of service areas of villages in a $k = 7$ hierarchy 370

22.12 Hamlets, villages and towns in a 'fixed-k' hierarchy $k = 4$ (Christaller's traffic principle) 370

22.13 'Fixed-k' hierarchy of $k = 7$ developed to third order of settlement (market towns) (Christaller's administrative principle) 370

22.14 Alternative nesting principles in the Christaller 'fixed-k' landscape 371

22.15 Löschian landscape developed by the rotation of a number of k-system hexagons, developing 'city-rich' and 'city-poor' sectors 375

22.16 Locational features of the study area, Snohomish County, Washington 378

22.17 Classes of central places in Snohomish County 382

22.18 Rural service centres in south-western Wisconsin 386

22.19 Rural service centres in southern England 387

23.1 Eastern Europe before the nineteenth century 396

23.2 Political divisions of Eastern Europe at the time of the Third Crusade 396

23.3 Political divisions of Eastern Europe at the accession of Charles V 397

23.4 Continental and arctic drainage (equal-area projection) 399

23.5 The natural seats of power 403

23.6 The Heartland concept 410

23.7 The Heartland concept 410

23.8 The airman's view: de Seversky (1950) 411

23.9 Rural frontier zones in the Alpes Maritimes 420

23.10 The Limburg appendix 421

26.1 Genealogy of Tiv lineages 472

26.2 Map and genealogy of Tiv political divisions 474

27.1 Inputs–political system or processes–outputs 492

31.1 Mean maxima in proximity series 597

31.2 Level of tension and nervousness, showing the self-reports on 'tension and nervousness' for 137 subjects in the proximity experiments 601

31.3 Predicted and obtained behaviour in voice feedback 605

33.1–6 Population change and variability, six types 629

33.7 Regional types of variability change 630

33.8 Malaria regions, 1938 638

33.9 Increase of population, 1951–61 639

33.10 Levels of economic development 642

Chart 1 World according to population and birth rates by regions 617

Chart 2 Growth in world population 618

Chart 3 Age structure of population in selected regions 618

Chart 4 Trends of births, death and natural increase rates in selected countries 619

Chart 5 World population by major areas and regions 619

LIST OF SOURCES

Chapter

1 J. Z. Young, in J. Z. Young and Tom Margerison (eds), *From Molecules to Man* (Thames & Hudson, 1969) pp. 88–9.

E. D. Erb and D. Hooker, *The Psychology of the Emerging Self* (F. A. Davis, 1967) pp. 1–13.

2 A. P. d'Entrèves, *The Notion of the State* (Clarendon Press, Oxford, 1967) pp. 153–68, 201–21.

3 A. Smith, *The Wealth of Nations* (1776) (Everyman ed.) pp. 4–7, 400.

R. A. Radford, 'The Economic Organisation of a Prisoner of War Camp', *Economica* (Nov 1945).

4 H. C. Brookfield, 'The Chimbu: A Highland People in New Guinea', in S. R. Eyre and G. R. Jones (eds), *Geography as Human Ecology* (Edward Arnold, 1966).

'The Money that Grows on Trees', *Australian Geographical Studies*, vol. 6, no. 2 (Oct 1968) 97–119.

5 W. F. Whyte, *Street Corner Society*, 2nd ed. (University of Chicago Press, 1955) pp. xv–xvi, 9–13, 25–35, 255–63.

L. Marshall, 'Sharing, Talking and Giving: Relief of Social Tensions among !Kung Bushmen', *Africa* (Mar 1961).

6 J. Dollard and N. E. Miller, *Personality and Psychotherapy* (McGraw-Hill, 1950) pp. 132–54.

C. S. Ford and F. A. Beach, *Patterns of Sexual Behaviour* (Eyre & Spottiswoode, 1952) pp. 167–98.

7 W. F. Ogburn, 'The Wolf Boy of Agra', *American Journal of Sociology*, vol. 64, no. 5 (University of Chicago Press, 1959) 449–54.

B. Bettelheim, 'Feral Children and Autistic Children', *American Journal of Sociology*, vol. 64, no. 5 (University of Chicago Press, 1959) 455–67.

Chapter

8 H. C. Kelman, in M. Jahoda and N. Warren (eds), *Attitudes* (Penguin Books, 1966) pp. 151–61.

B. Bettelheim and M. Janowitz, 'Ethnic Tolerance', *American Journal of Sociology*, vol. 55 (University of Chicago Press, 1949) 137–45.

I. Sarnoff, D. Katz and C. McClintock, D. Katz *et al.* (eds), *Public Opinion and Propaganda* (Holt, Rinehart & Winston, 1954) pp. 121–8.

9 E. F. Vogel, in M. F. Nimkoff (ed.), *Comparative Family Systems* (Houghton Mifflin, 1965) pp. 287–99.

D. J. Stenning, *Savannah Nomads* (Oxford University Press, for International African Institute, 1959) pp. 100–30.

H. Gavron, *The Captive Wife* (Routledge & Kegan Paul, 1966) pp. 34–41.

10 A. Marshall, *Principles of Economics*, 7th ed. (Macmillan, 1920) pp. 63, 92–9.

J. K. Galbraith, *The Affluent Society* (Penguin Books, 1962) pp. 131–7.

11 R. Turvey, 'How Rational Economic Decisions are Made', *Lloyds Bank Review* (Jan 1968) pp. 39–45.

J. A. Schumpeter, *Capitalism, Socialism and Democracy* (Allen & Unwin, 1947) pp. 83–6.

12 E. H. Phelps Brown and J. Wiseman, *A Course in Applied Economics* (Pitman, 1960) pp. 26–43.

A. Nove, *The Soviet Economy* (Allen & Unwin, 1968) pp. 171–96.

13 S. Matsushima, 'Labour–Management Relations in Japan', in Paul Halmos (ed.), *Japanese Sociology Studies*, Sociological Review Monograph 10 (1966) pp. 69–81.

D. Roy, 'Efficiency and "the Fix": Informal Intergroup Relations in a Piecework Machine Shop', *American Journal of Sociology*, vol. 60 (University of Chicago Press, 1955) 255–66.

Chapter

14 J. Robinson, *Economic Philosophy*, New Thinkers Library (C. A. Watts, 1962) pp. 89–98.

H. G. Johnson, *Money, Trade and Economic Growth* (Allen & Unwin, 1962) pp. 107–16.

15 J. R. Hicks, *Critical Essays in Monetary Theory* (Oxford University Press, 1967) pp. 155–73.

16 R. J. Nicholson, 'The Distribution of Personal Income', *Lloyds Bank Review*, no. 83 (Jan 1967) pp. 11–21.

17 M. N. Srinivas, *Religion and Society among the Coorgs of South India* (Clarendon Press, Oxford, 1952) pp. 25–31.

18 T. H. Pear, *English Social Differences* (Allen & Unwin, 1955) pp. 58–83.

19 J. H. von Thünen, *Isolated State*, trans. C. M. Wartenburg, ed. Peter Hall (Pergamon, 1966) p. 7.

A. Grötewald, 'Von Thünen in Retrospect', *Economic Geography*, vol. 35 (1959) 348–53.

J. W. Alexander, *Economic Geography* (Prentice-Hall, 1963) pp. 473–6.

J. J. Hidore, 'The Relationship between Cash-Grain Farming and Landforms', *Economic Geography*, vol. 39 (1963) 84–9.

J. Wolpert, 'The Decision Process in Spatial Context', *Annals of the Association of American Geographers*, vol. 54 (1964) 537–58.

J. R. Peet, 'The Spatial Expansion of Commercial Agriculture', *Economic Geography*, vol. 45 (1969) 283–301.

20 F. E. I. Hamilton, in R. J. Chorley and P. Haggett (eds), *Models in Geography* (Methuen, 1967) p. 371.

P. Haggett, *Location Analysis in Human Geography* (Edward Arnold, 1965) pp. 143–5.

W. Smith, 'The Location of Industry', *Transactions of the Institute of British Geographers*, vol. 21 (1955) 1–18.

J. G. March and M. A. Simon, *Organisations* (Wiley, 1958) pp. 140–1.

Chapter

M. L. Greenhut and M. R. Colberg, *Factors in the Location of Florida Industries* (Florida State University Press, 1962) p. 44.

F. E. I. Hamilton, 'Location Factors in the Yugoslav Iron and Steel Industry', *Economic Geography*, vol. 40 (1964) 57–61.

J. R. Tarrant, 'Industrial Development in Ireland', *Geography*, vol. 52 (1967) 404–7.

L. Hamill, 'A Test of Weberian Location Theory', *Abstract of Papers, 20th International Geographical Union* (London, 1964) p. 234.

T. E. Macmillan, 'Why Manufacturers Choose Plant Locations *v.* Determinants of Plant Location', *Land Economics*, vol. 41 (3) (1965) 239–46.

21 W. R. Seyfried, 'The Centrality of Urban Land Values', *Land Economics*, vol. 39 (1963) 275–6.

B. J. L. Berry, R. J. Tennant, B. J. Garner and J. W. Simmons, *Commercial Structure and Commercial Blight* (Department of Geography, University of Chicago, 1963), Research Paper No. 85, p. 14.

H. Hoyt, *One Hundred Years of Land Values* (University of Chicago Press, 1939) pp. 297–8.

M. H. Yeates, 'Some Factors Affecting the Spatial Distribution of Chicago Land Values 1910–1960', *Economic Geography*, vol. 41 (1965) 57–68.

R. E. Parke and E. W. Burgess, *The City* (University of Chicago Press, 1925) pp. 50–1, 54–6.

B. J. L. Berry *et al.*, *Studies of Highway Development and Geographical Change* (Prentice-Hall, 1959) p. 64.

G. W. Hartmann, 'The Central Business District: A Study in Urban Geography', *Economic Geography*, vol. 26 (1950) 238–9.

P. J. Smith, 'Calgary: A Study in Urban Patterns', *Economic Geography*, vol. 38 (1962) 321, 323–9.

C. D. Harris and E. L. Ullman, 'The Nature of Cities', *Annals of the American Academy of Political and Social Science*, vol. 242 (1945) 14–17.

Chapter

P. Haggett, *Location Analysis in Human Geography* (Edward Arnold, 1965) pp. 180–1.

22 P. Ambrose, *Analytical Human Geography* (Longmans, 1969) pp. 121–30.

J. A. Everson and B. P. FitzGerald, *Settlement Patterns* (Longmans, 1969) pp. 102–9.

B. J. L. Berry and A. Pred, *Bibliography of Theory and Application: Central Place Studies* (Regional Science Research Institute, Philadelphia, 1961) series 1, pp. 15–17.

E. Ullman, 'A Theory of Location for Cities', *American Journal of Sociology*, vol. 46 (1941) 853–64.

B. J. L. Berry and W. L. Garrison, 'The Functional Bases of the Central Place Hierarchy', *Economic Geography*, vol. 34 (1958) 145–54.

H. M. Mayer and C. F. Kohn, *Readings in Urban Geography* (University of Chicago Press, 1965) p. 184.

J. E. Brush and H. E. Bracey, 'Rural Service Centers in Southwestern Wisconsin and Southern England', *Geographical Review*, vol. 45 (1955) 559–69.

W. Bunge, *Theoretical Geography* (Lund Studies in Geography, 1966) p. 133.

23 H. J. Mackinder, 'The Geographic Pivot of History', *Geographical Journal*, vol. 23 (1904) 421–44.

H. de Blij, *Systematic Political Geography* (Wiley, 1967) pp. 130–7.

J. W. House, in C. A. Fisher, *Essays in Political Geography* (Methuen, 1968) pp. 327–44.

24 R. A. Dahl, *Modern Political Analysis* (Prentice-Hall, 1963) pp. 25–31.

G. Sartori, 'Constitutionalism: A Preliminary Discussion', *American Political Science Review* (Dec 1962) pp. 853–64.

25 S. Neumann, *Modern Political Parties* (University of Chicago Press, 1955).

Chapter

Gabriel A. Almond, 'A Comparative Study of Interest Groups and the Political Process', *American Political Science Review*, vol. 52 (1958) 270–2.

26 P. Bohannan, in J. L. Gibbs (ed.), *Peoples of Africa* (Holt, Rinehart & Winston, 1965) pp. 523–37.

S. MacCallum, in P. Bohannan (ed.), *Law and Warfare* (Natural History Press, New York, 1967) pp. 291–9.

L. Mann, 'Queue Culture: The Waiting Line as a Social System', *American Journal of Sociology*, vol. 75 (1969) 345–58.

27 D. Easton, 'An Approach to the Analysis of Political Systems', *World Politics* (Apr 1957) pp. 383–400.

R. H. Tawney, *Religion and the Rise of Capitalism*, 1st ed. (Penguin Books, 1938) pp. 228–39, 269–78.

M. Weiner, 'Struggle against Power: Notes on Indian Political Behaviour', *World Politics* (Apr 1956) pp. 392–404.

L. Lipsitz, 'Work Life and Political Attitudes: A Study of Manual Workers', *American Political Science Review*, vol. 58 (1964) 951–62.

28 R. Michels, *Political Parties* (Constable, 1915).

R. Frankenburg, *Village on the Border* (Cohen & West, 1957) pp. 100–47.

29 P. Cohen, *Modern Social Theory* (Heinemann, 1968) pp. 18–33.

30 R. Merton, *Social Theory and Social Structure* (Free Press, Glencoe, Ill., 1957) pp. 71–82.

L. Coser, *The Functions of Social Conflict* (Free Press, Glencoe, Ill., 1956) pp. 151–6.

31 S. E. Asch, in G. E. Swanson *et al.* (eds), *Readings in Social Psychology* (Holt, Rinehart & Winston, 1952) pp. 393–401.

S. Milgram, 'Some Conditions of Obedience and Disobedience to Authority', *Human Relations*, vol. 18 (1965) 57–74.

T. H. Newcomb, in G. E. Swanson *et al.* (eds), *Readings in Social Psychology* (Holt, Rinehart & Winston, 1952) pp. 215–25.

Chapter

32 N. B. Ryder, 'The Reproductive Renaissance North of the Rio Grande', *Annals of the American Academy of Political and Social Science* (1958) 18–24.

Finance and Development (U.N. Publication) vol. 6, no. 4 (Dec 1969) 48 ff.

33 A. Geddes, 'The Social and Psychological Significance of Variability in Population Change', *Human Relations*, vol. 1 (1947–8) 181–205.

A. T. A. Learmonth, 'India's Population Geography in an Era of Change', amended from 'Selected Aspects of India's Population Geography', *Australian Journal of Politics and History*, vol. 12, no. 2 (1966).

34 J. R. Malthus, *An Essay on the Principles of Population* (1798) (Everyman ed.) pp. 5–11, 18–19.

Chapter

H. Myint, *The Economics of the Developing Countries* (Hutchinson University Library, 1964) pp. 29–36.

W. A. Lewis, *A Theory of Economic Growth* (Allen & Unwin, 1955) pp. 319–30.

35 M. Weiner, 'Urbanisation and Political Protest', *Civilisations*, vol. 17, no. 1/2 (1967) 44–50.

R. Freedman, 'Norms for Family Size in Underdeveloped Areas', *Proceedings of the Royal Society*, series B, vol. 159 (1963) 220–45.

36 E. Katz, 'The Two-step Flow of Communication', *Public Opinion Quarterly*, vol. 21 (1957) 61–78.

B. Berelson, 'On Family Planning Communication', *Demography*, vol. 1, no. 1 (1964) 94–105.

ACKNOWLEDGEMENTS

The Association of American Geographers for J. W. Alexander's Figs 26–5 to 26–8, and J. Wolpert, 'The Decision Process in Spatial Context' *Annals* (vol. 54, 1964). Asia Publishing House and Penguin Books Ltd for 'The Caste System in India', from *Religion and Society among the Coorgs of South India* by M. N. Srinivas. *Australian Geographical Studies* for 'The Money that Grows on Trees', vol. 6, no. 2 (Oct 1968) pp. 97–119, by H. C. Brookfield. *Journal of Abnormal Social Psychology* (vol. 49, 1954) for extract from 'The Motivational Basis of Attitude Change', from *Public Opinion and Propaganda*, ed. Daniel Katz *et al.* (Holt, Rinehart & Winston, 1954). The *Annals of the American Academy of Political and Social Science*, E. L. Ullman and C. D. Harris for 'The Nature of Cities' (vol. 242, 1945), and 'The Reproductive Renaissance North of the Rio Grande' (vol. 316, 1958) by N. B. Ryder. George Allen & Unwin Ltd and F. A. Praeger, Inc., for extract from *Capitalism, Socialism and Democracy* by J. A. Schumpeter. George Allen & Unwin Ltd for extracts from *The Soviet Economy* by A. Nove; *Money, Trade and Economic Growth* by Harry G. Johnson; *English Social Differences* by T. H. Pear, chap. 2, 'Psychological Aspects of English Social Stratification'; and *A Theory of Economic Growth* by W. A. Lewis (Richard D. Irwin, Homewood, Ill.). American Political Science Association for 'Constitutionalism: A Preliminary Discussion' by G. Sartori; 'A Comparative Study of Interest Groups and the Political Process' by Gabriel A. Almond; and 'Work Life and Political Attitudes: A Study of Manual Workers' by L. Lipsitz. Edward Arnold (Publishers) Ltd for extracts from P. Haggett, *Location Analysis in Human Geography*, and Figs 6.17 and 6.18 from above; and 'The Chimbu: A Highland People in New Guinea' by H. C. Brookfield, from *Geography as Human Ecology*, ed. S. R. Eyre and G. R. Jones. Marshall R. Colberg and M. L. Greenhut for extract from *Factors in the Location of Florida Industry*. Carnegie Press for 'Effects of Group Pressure upon the Modification and Distortion of Judgments' by S. E. Asch, from *Readings in Social Psychology*, ed. Guy E. Swanson *et al.* Clarendon Press, Oxford, for extracts from Part 3 of *The Notion of the State* (1967) by A. P. d'Entrèves, and chap. 9 of *Critical Essays in Monetary Theory* (1967) by John Hicks. *Economic Geography*, Clark University, for extracts from 'The Functional Bases of the Central Place Hierarchy' by B. J. L. Berry and W. L. Garrison (vol. 34, 1958); 'Location Factors in the Yugoslav Iron and Steel Industry' by F. E. I. Hamilton (vol. 40, 1964); 'Some Factors Affecting the Spatial Distribution of Chicago Land Values' by M. H. Yeates (vol. 41, 1965); 'Calgary: A Study in Urban Patterns' by P. J. Smith (vol. 38, 1962); 'The Central Business District: A Study in Urban Geography' by G. W. Hartmann (vol. 26, 1950); 'The Spatial Expansion of Commercial Agriculture' by J. R. Peet (vol. 45, 1960); 'The Relationship between Cash-Grain Farming and Landforms' by J. J. Hidore (vol. 39, 1963); 'Von Thünen in Retrospect' by Andreas Grötewald (vol. 35, 1959); diagrams from 'Selected Aspects of Uniform and Nodal Regions' by Alexander, Brown and Dahlberg (vol. 34, 1958). F. A. Davis Company for extract from *The Psychology of the Emerging Self* by Everett D. Erb and D. Hooker. Doubleday & Company, Inc., for 'Dispute Settlement in an American Supermarket' by S. MacCallum, from *Law and Warfare*, ed. Paul Bohannan. J. M. Dent & Sons Ltd for the extracts 'The Division of Labour' and 'The "Invisible Hand" ' from *The Wealth of Nations* by Adam Smith, and from *An Essay on the Principle of Population* by J. R. Malthus. Dover Publications Inc. for extract from *Political Parties* (1951) by R. Michels, chap. 6. Department of Geography, University of Chicago, for Research Paper No. 85, *Commercial Structure and Commercial Blight*, p. 14, by B. J. L. Berry, R. J. Tennant, B. T. Garner and J. W. Simmons. *Geographical Journal* for H. J. Mackinder, 'The Geographical Pivot of History' (vol. 23, 1904). Institute of British Geographers for 'The Location of Industry' (*Transactions*, vol. 21, 1955) by W. Smith. J. E. Brush and H. E. Bracey, 'Rural Service Centers in Southwestern Wisconsin and Southern England' reprinted from the *Geographical Review* (vol. 45, 1955), © by the American Geographical Society of New York. The Geographical Association for 'Industrial Development in Ireland' (*Geography*, vol. 52 (1967) pp. 404–7) by J. R. Tarrant. Hamish Hamilton, London, and Houghton Mifflin, Boston, for extract from *The Affluent Society*,

copyright © 1958, 1969 by John Kenneth Galbraith. Wm. Heinemann Educational Books Ltd, and Basic Books, Inc., New York, for 'The Problem of Order' from *Modern Social Theory* (1968) by Percy Cohen, chap. 2. Houghton Mifflin Company for 'The Japanese Family', by Ezra F. Vogel, from *Comparative Family Systems*, ed. M. F. Nimkoff (Boston, 1965) chap. 14. Hutchinson Publishing Group and F. Praeger, Inc., New York, for extract from *The Economics of the Developing Countries* (1964) by H. Myint. Institut International des Civilisations Différentes for extract from *Civilisations* (Brussels), vol. 17 (1967) no. 1/2, pp. 1–7, by Myron Weiner. International African Institute for extract from *Savannah Nomads* by D. J. Stenning, and 'Sharing, Talking and Giving; Relief of Social Tensions among !Kung Bushmen' by Lorna Marshall. The London School of Economics and Political Science for 'The Economic Organisation of a P.O.W. Camp', from *Economica* (Nov 1945) by R. A. Radford. Lloyds Bank Review Ltd for 'The Distribution of Personal Income' (no. 83, Jan 1967) by R. J. Nicholson, and 'How Rational Economic Decisions are Made' (Jan 1968) by R. Turvey. *Land Economics* and W. R. Seyfried for extract from 'The Centrality of Urban Land Values' (vol. 39, 1963), and T. E. Macmillan, Jr, for 'Why Manufacturers Choose Plant Locations *v.* Determinants of Plant Location' (vol. 41(3), 1965). Longman Group Ltd for extracts from *Analytical Human Geography* by P. J. Ambrose, and *Settlement Patterns* by J. A. Everson and B. P. FitzGerald, both from the 'Concepts in Geography' series. Holt, Rinehart & Winston, Inc., for 'Attitude Development as a Function of Reference Groups' by T. M. Newcomb, from *Readings in Social Psychology* (1952), ed. G. E. Swanson *et al.*, and extracts from 'The Tiv of Nigeria' by Paul Bohannan, in *Peoples of Africa*, ed. J. L. Gibbs, Jr, © 1965 Holt, Rinehart & Winston Inc. Lund Studies in Geography, Royal University of Lund, for extract from *Theoretical Geography* by W. Bunge. Macmillan Company, New York, for extract from *Social Theory and Social Structure* by Robert Merton. Extract from *Personality and Psychotherapy* by J. Dollard and N. E. Miller, copyright 1950 by McGraw-Hill Inc., used with permission of the McGraw-Hill Book Company. Methuen Educational Ltd for 'A Local Perspective on Boundaries and the Frontier Zone' by J. W. House from *Essays in Political Geography* (1968) chap. 18, ed. C. A. Fisher; and F. E. I. Hamilton, in *Models in Geography* (1967), ed. R. J. Chorley and P. Haggett. Laurence Pollinger Ltd, and Harper & Row, Inc., New York, for 'Development in the Individual', from *Patterns of Sexual Behaviour* by C. Ford and F. A. Beach. Penguin Books Ltd for extract from *Religion and the Rise of Capitalism* by R. H. Tawney. Pergamon Press Ltd for extract from *Von Thünen's Isolated State* (1966), trans. C. M. Wartenburg, ed. Peter Hall. Sir Isaac Pitman and Sons Ltd for extract from *A Course in Applied Economics* (1960) by E. H. Phelps Brown and J. Wiseman. Plenum Publishing Company Ltd for extract from 'Some Conditions of Obedience and Disobedience' by Stanley Milgram, and 'The Social and Psychological Significance of Variability in Population Change' by A. Geddes. Princeton University Press for 'An Approach to the Analysis of Political Systems' by D. Easton (*World Politics*, Apr 1957), and 'Struggle against Power: Notes on Indian Political Behaviour' by M. Weiner (*World Politics*, 8, Apr 1956), © 1957. *Public Opinion Quarterly* for extract from 'The Two-Step Flow of Communication' by Elihu Katz (vol. 21, 1957) and 'Three Processes of Social Influence' by H. C. Helman (vol. 25, 1961). Prentice-Hall, Inc., for John W. Alexander, *Economic Geography*, © 1963, and R. A. Dahl, *Modern Political Analysis*, © 1963, reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J. Routledge & Kegan Paul Ltd for 'The Politics of Recreation', from *Village on the Border* by R. Frankenburg (Cohen & West, 1957). Routledge and Kegan Paul Ltd and Humanities Press for extract from *The Captive Wife* (1966) by Hannah Gavron. Routledge & Kegan Paul Ltd and Free Press, Inc., for extract from *The Functions of Social Conflict* (1956) by Lewis Coser. Regional Science Research Institute, Philadelphia, for extract from *Bibliography of Theory and Application: Central Place Studies*, 1 (1961) by B. J. L. Berry and A. Pred. The Royal Society for extract from 'Norms for Family Size in Underdeveloped Areas' by Ronald Freedman (*Proceedings*, vol. 159, 1963), and 'A Test of Weberian Location Theory' by Louis Hamill (*Abstract of Papers, 20th International Geographical Union*, London, 1964). Sociological Review Monograph for Shizuo Matsushima, 'Labour–Management Relations in Japan', No. 10 in *Japanese Sociology Studies*, ed. Paul Halmos (1966). Thames & Hudson Ltd and Crown Publishers, New York, for extract from *From Molecules to Man* by J. Z. Young and T. Margerison. Edward Ullman for reprint of 'A Theory of Location for Cities', from *American Journal of Sociology*

(vol. 46, 1941), also to appear in *Geography as Spatial Interactions: Studies in Regional Development* (University of Washington Press, 1971), by Edward Ullman. The University of Chicago Press for 'On Family Planning Communications' by B. Berelson (*Demography*, vol. 1, no. 1, 1964). J. Wiley and Sons, Inc., for extract from *Organisations* (1958) by J. G. March and M. A. Simon, and 'The Heartland Debate' from H. de Blij, *Systematic Political Geography* (1967). C. A. Watts and Company for chap. 4 of *Economic Philosophy* (1962) by Joan Robinson. University of Washington Press for *Studies of Highway Development and Geographical Change* (1959) by B. J. L. Berry *et al.* (1959). University of Chicago Press for extracts from *Street Corner Society*, 2nd ed. (1955) by W. F. Whyte; *The City* (1925) chap. 2, by R. E. Parke and E. W. Burgess; *Readings in Urban Geography* (1965) by H. M. Mayer and C. F. Kohn; 'Towards a Comparative Study of Political Parties', from *Modern Political Parties* (1955) by S. Neumann; *One Hundred Years of Land Values* (1939) by H. Hoyt; also (taken from *American Journal of Sociology*) 'The Wolf Boy of Agra' by W. F. Ogburn (vol. 64, no. 5, 1959); 'Feral Children and Autistic Children' by B. Bettelheim (ibid.); 'Ethnic Tolerance' (vol. 55, 1949) by B. Bettelheim and M. Janowitz; 'Efficiency and "the Fix": Informal Intergroup Relations in a Piecework Machine Shop' by D. Roy (vol. 60, 1955); 'Queue Culture: The Waiting Line as a Social System' (vol. 75, 1969) by Leon Mann.

INTRODUCTION

This anthology of readings in the social sciences has been compiled by members of the Social Sciences Course Team of the Open University, as part of their Foundation Course 'Understanding Society'. It ranges across the fields of economics, geography, government, psychology and sociology, and includes both new and classic material in each discipline.

The reader is mainly designed for students of the Open University who are taking the 'Understanding Society' course; for them it represents essential reading. But it should be of considerable interest to other students of the social sciences, containing as it does a wide variety of extracts from books and periodicals, drawn from both British and overseas sources across each of the disciplines.

The Open University is concerned primarily with adult students in full-time employment or working in the home. It was set up to provide higher educational opportunities for those who, for one reason or another, were unable to benefit from a conventional university education, and who now wish to extend their studies. In the words of the original Planning Committee report: 'We took it as axiomatic that no formal academic qualifications would be required for registration as a student. Anyone could try his or her hand, and only failure to progress adequately would be a bar to continuation of studies.'

Students are not confined to one or two subjects; their work programmes are designed to embrace studies over a wide range of subjects and across several disciplines. Students are able to study both arts and science subjects and there are as few restrictions as possible on choice. Maximum flexibility and an interdisciplinary approach are the characteristics of Open University courses.

Each course is built up around correspondence 'packages', sent to the students by post and calling for the return of assignments by the students; integrated with these correspondence packages are series of television and radio programmes, short summer or week-end courses and a regionally organised tutorial and counselling system.

The correspondence courses use modern learning methods, including programmed learning techniques where appropriate. The broadcast programmes also pioneer new paths; they do not reproduce conventional university lectures but experiment with all the resources of the media.

The courses are, therefore, carefully planned and integrated instructional systems, created by Course Teams which are made up of academic staff, educational technologists, broadcasting and media specialists. This reader is an essential component of the 'Understanding Society' course, and biographical notes on its editors can be found at the end of the volume.

'Understanding Society' is a course very much in keeping with the spirit of an 'open' university; it is an integrated and interdisciplinary approach to the social sciences. It considers five different views of man – the economic, sociological, psychological, political and geographic – in order to highlight these various facets of his life in society.

The course is divided into three parts which ask three fundamental questions: why people live in societies; how people live in societies; and what kind of problems they face. The first part looks at such questions as: What does man gain from being under government? What are the economic advantages man secures through living in society? What biological, emotional and social needs does he seek to satisfy?

The second and major section of the course is focused on how men live in societies. This is not simply a descriptive account of the various social, economic and political systems which man has developed or of the environments in which he has developed them. It also attempts to analyse, explain and, within limits, predict man's behaviour, drawing when appropriate on scientific skills and methods. Each discipline thus presents a range of its insights, and many examples are drawn from societies contrasted in complexity.

The final section of the course takes, by way of example, one of the major problems that faces the world today: the so-called 'population explosion'. This is a multi-faceted problem in the study of which all the social sciences have a contribution to make. As Barbara Wootton has remarked, it may fall to the social sciences

'to bridge the chasm between the scientific and literary worlds, which is today so widely regarded as a threat to the integrity of contemporary culture. While our subject-matter links us to the humanities, our methods approximate ever more closely to those of the natural sciences.'

This anthology adopts the same pattern as the course itself and includes materials which, in the opinion of its editors, have something particularly relevant or illuminating to say about the three fundamental questions posed. It should provide a foundation for further work and reading in the social sciences; in what it has to say about man's role in society it should also be of interest to anyone concerned with the problems of the contemporary world.

PART ONE

Why People Live in Societies

If superiority is measured in terms of cultural, scientific and technological achievement, then man is vastly superior to any other creature living on this planet. Because of this massive superiority, Charles Darwin originally had great difficulty in persuading the leading scientists and churchmen of his day that man evolved, like every other animal, from more primitive life-forms. Nevertheless there are good grounds for believing that Darwin was right. And a great deal of man's day-to-day behaviour still testifies to his animal nature and ancestry.

In the first extract, an international authority (Professor J. Z. Young of University College, London) briefly reviews what is known about man's evolutionary history.

In the second extract, from 'The Psychology of the Emerging Self' by Everett D. Erb and Douglas Hooker, our main focus of attention shifts from the biological origins to the behaviour *of man – and to some of the many problems arising out of man's attempts to explain his own behaviour.*

The authors comment on the difficulty of giving conclusive explanations of human behaviour, and on the inadequacies of 'common-sense' explanations. Finally, they discuss some of the basic needs – both physical and psychological – that prompt men to act as they do.

FROM MOLECULE TO MAN

J. Z. Young

THE beginning of life as we know it probably took place within the 4000 million years or thereabouts for which the earth has existed. It is possible, of course, that life may have reached the planet from some other source. The discovery in meteorites of compounds that are usually associated with life, such as hydrocarbons and even purines and pyrimidines, has suggested to some people that life exists or has existed elsewhere than on the earth. But this evidence can be interpreted differently if the bodies in our immediate neighbourhood were derived from the earth, for example, as a result of tidal disturbances of the crust, following a very near approach of the moon. According to this view the continents may have dropped from the moon to the earth, perhaps as a single mass which later fragmented and drifted apart. Pieces of the original crust may have been thrown into space, perhaps carrying life, or the remains of it, with them.

That such speculations are plausible shows the difficulties and uncertainties which must be faced. Nevertheless studies of the origin of life have assumed some degree of exactness in recent years. We can begin to see three possible stages by which the process may have taken place. First, there must have been the formation of organic molecules themselves. Secondly, these have somehow become collected or aggregated to form an early 'protoplasmic' material, with self-maintaining and self-reproducing properties. Thirdly, only perhaps at a relatively late stage, this material evolved into the variety of plants and animals as we know them. Of these three stages the second is much the most difficult to understand. There is actual experimental evidence about how the other two stages occurred. They can indeed even be repeated under controlled laboratory conditions today. But we do not know how the early organic molecules came together to make self-maintaining systems.

The most generally accepted view of the origin of organic molecules is that of Haldane and Oparin, that they were formed from the gases in the earth's original atmosphere. An important aspect of the problem is that all living things consist in the main of the same few elements. The great bulk of organic matter is composed of carbon, hydrogen, oxygen and nitrogen. Most organisms also contain smaller amounts of sulphur, phosphorus, chlorine, sodium, potassium, calcium, magnesium and iron. A considerable number of other elements also occur in small quantity, but not in all organisms. These 'trace elements', for example iodine, copper and even zinc, gold or vanadium, may indeed form necessary parts of the enzyme system. However, it is important to emphasise that life as we know it consists essentially of various combinations of the four main elements, carbon, hydrogen, oxygen and nitrogen. This introduces us at once to the fact that *life involves a selection among a number of possible events*. This feature of selection of certain

combinations from a set we shall find repeatedly throughout our discussion of the origin and evolution of life.

It will be noted that the elements of living matter are not by any means the commonest of those on the earth. Aluminium and silicon, though they occur in some plants, are not prominent in living things. We have then first to inquire by what process there was produced the selection of elements that has persisted ever since.

THE BASIC MOLECULES OF LIFE

The composition of the gases of the early atmosphere is not known for certain, but perhaps included water (in the form of steam), carbon dioxide, carbon monoxide, the chloride, fluoride and sulphide of hydrogen, nitrogen, ammonia and methane. There was probably no free oxygen. From these gases, at sufficiently high temperature, simple organic molecules could have been formed with the energy available from short-wave (ultra-violet) radiation. This would have reached the earth at much greater intensity than it does today, when it is screened off by oxygen and ozone. Some very impressive experiments by Miller and others have shown that such a combination can take place. Miller passed a spark discharge (instead of radiation) through mixtures of gases such as hydrogen, methane, ammonia and water vapour, simulating the primæval atmosphere. Analysis of the contents of the vessel after the spark had passed showed appreciable amounts of such organic substances as formic, acetic and propionic acids, urea, lactic and succinic acids, many amino-acids, the purines adenine and guanine, the sugars ribose and deoxyribose, adenosine and AMP, ADP and ATP. These are all substances common in living things but not known to occur elsewhere in nature.

There is therefore no longer any doubt that the first stage of the formation of life can occur under the influence of entirely 'natural' conditions such as may well have existed early in the earth's history. But these substances are all relatively simple combinations. Their molecules are made up at most of a few tens of atoms. The characteristic living molecules include the nucleic acids, sugars, fatty substances and proteins, each of which may be made up of millions of atoms. How could the early organic compounds have come to make these combinations and how is it ensured that they persist or are repeatedly reproduced?

Once again the problem is seen to be not quite so formidable when we find that the secret of the ordering of the atoms in large molecules is the same in all organisms from viruses to man. It depends upon the order of the bases in one particular type of organic molecule, the nucleic acids, which have been described in detail. The centre of the problem at this point is to discover how these units may have been formed from the first simple molecules and how they may have become aggregated to make self-reproducing chains. This is indeed a very difficult problem, but the further question of how the nucleic acids came to be able to make proteins is really equally formidable.

Once again we see that life involves essentially certain selected combinations, out of many possible ones. The very early stages of biochemical evolution may have proceeded by chance interactions, only the more stable products being eventually selected. Here again we have help from the laboratory, since it has been shown by Oró that the repeating bases in the nucleic acid structure can be formed from the simple substances ammonia and cyanogen if sufficient energy is available.

It might seem at first sight that we must appeal to 'chance' processes for the formation of these large ordered organic molecules and that such events would be very improbable. But it cannot be excluded that life took its aspect from some ordered feature of the inorganic world. Bernal has called attention to the possibility that clays in the estuaries of early seas might have provided surfaces to which the early organic molecules became attached. Such a process of adsorption might have had the double effect of ordering the arrangement of the organic molecules to fit the chemical structure of the clays and of either activating or stabilising them by lining them up and directing them. Such effects of adsorption are commonly used by chemists in catalysis, enabling reactions between substances to proceed under conditions of temperature and pressure in which they would not interact without the catalyst. Moreover we know that the clays and related substances form a large variety of quite complex molecules, bauxites, quartzes and the like. It would be strangely paradoxical if the aluminium and silicon of which they are chiefly composed turned out to be the originators of the pattern of living matter, without entering into it themselves.

A particular feature of living molecules that has to be explained

is that where two symmetries are possible it is always either the right- or left-handed one that is found in living materials, whereas as a result of an ordinary chemical reaction both are produced together. Pasteur showed that pure left-handed samples can be prepared by picking out the crystals by hand – but this is, after all, the work of a living creature. Pirie has suggested the possibility that this strange handedness of life was due to the fact that the early organic molecules were formed on quartz crystals, which may be very long, thin and asymmetrical, with active hydroxyl (–OH) groups at the surface.

The formation of large organic molecules is therefore not quite so impossible to imagine as it would be if we appealed only to a 'random' process. On the one hand there is the fact that the critical materials, the nucleic acids, though immensely complicated, are composed of relatively simple units, repeating in various combinations. On the other hand we can imagine possible ways in which some ordering could be obtained from non-living materials. We must recognise, however, that the speculation about this stage is much vaguer than at the previous one. We have no laboratory experiments by which large molecules have been formed in this way, but it is possible to imagine that they could be made. It is not by any means certain that we are right to concentrate on the nucleic acids as the primary source of order at this stage of the formation of life. They certainly became so at some stage, but as we know them today they are only able to reproduce themselves in association with proteins, and it may be that we must look for the origin of a nucleoprotein aggregate rather than for either part of it alone.

THE UNEXPLAINED ESSENTIAL STEP

Studies and speculations of these types lead us to the point at which we can imagine that there has been formed what Pirie has called a probiotic soup. It is at this stage that there is perhaps the greatest uncertainty about the origin of life. For with such a soup we are still a long way from even the simplest self-maintaining organism in the sense that we usually understand that word. However, the position of the viruses, anomalous as it is, may help us to see how the gap was bridged. The viruses are self-replicating provided that they find themselves in the correct environment, which must include the tissues of some true, self-maintaining animal or plant. Perhaps the material of the probiotic soup was similarly able to replicate itself by the use of energy reaching it directly from the sun. The step that introduced organisms as we know them may have arisen as a response to decreased availability of such energy. Unfortunately it is not known whether there was a decrease in short-wave radiation from the sun *before* the activities of plants released the large quantities of screening oxygen.

The earliest organisms may indeed have developed gradually the means for using what energy was available, ending with chlorophyll, the basic source of energy supplies of the modern living world. The complex organic pigment molecules such as chlorophyll were another of the 'discoveries' that must have been made in the history of life. They play a central part, with other molecules, in the cycles of chemical change by which reactions are made to take place at the expense of the sun's energy but at temperatures and pressures far below those that would be needed in a laboratory.

It may have been the advent of these photosynthetic and respiratory cycles that introduced the characteristic living phenomenon of dynamic self-maintenance. So far we have been considering the probiotic material as if it could have existed in a state of equilibrium. A virus can indeed do this at the present day. A solution, for example, of tobacco-mosaic virus is a chemical as stable as any other at ordinary temperatures. A bottle-full of virus does not need to respire to remain alive. True organisms, on the other hand, are not in equilibrium but maintain a steady state of interchange with their surroundings. This taking in of supplies of raw materials and energy and later returning them to the environment is indeed the characteristic by which we recognise animals and plants as being 'alive'. Only when the interchange ceases is a creature fully 'dead'.

It may be, of course, that such continual replacement is in some way a condition for the persistence of the larger macromolecules. If so, only by the elaborate cycles of photosynthesis and respiration would the macromolecules survive the nights. As in the previous considerations about the origin of organic molecules, it seems at first sight improbable that a dynamically self-maintaining system would arise by the operation of ordinary physical and chemical laws. Some special conditions must indeed have been present to 'select' those combinations that ensured continuity.

Perhaps the special feature was provided by the very nature of the carbon compounds that had appeared. Because of its four available bonds, carbon is able to form whole series of molecules varying slightly from each other. Slight variations of conditions may have produced crops of molecules of different sorts and some combinations of these possessed the powers of photosynthesis and respiration that enabled them to survive as dynamic living systems. The operation of selection among large sets of entities with slight differences is not the sort of 'mechanism' that seems at first to the 'engineering' type of mind to be likely to produce results of great precision and complexity. Yet in fact it is by random methods of this sort that the most exact tools are produced. Perfectly spherical or flat mirrors are produced by engines that allow pieces of rough material to wear against each other with movements as nearly random as possible. It is essential to the understanding of natural processes to realise the fact that random forces operating over long periods may produce unique results.

Of course, to assert that dynamic self-maintaining systems might have arisen in this way is not to prove that they can do so or did so in the past. This will presumably only be achieved by laboratory methods such as those that have thrown so much light on the origin of organic compounds. It would indeed be an optimistic biochemist who assembled materials and left them with the appropriate energy source in the hope that living organisms would appear.

SEEKING THE BLUEPRINTS OF EARLY LIFE

However, once again the apparently fantastically difficult problem is made a little easier when we look at the specification of what would have to be produced to constitute a simple organism. Not only do all organisms contain very much the same compounds, but they have them organised in very much the same ways. Electron micrographs show how much of cell structure is built up of systems of membranes. Indeed, it is plausible to consider the cell as a sort of sponge composed of channels of infolded surface membrane. The structure of cell membranes is still not fully understood, but we know that they contain the special type of fatty substance known as phosphatides, linked with proteins. Moreover membranes, wherever they are found, upon or within the cell, show the same three-layered structure when appropriately prepared for the electron microscope. Even more interestingly it is found that such membranes form spontaneously whenever phosphatides are present in appropriate conditions. We are therefore not very far from being able to understand how this fundamental component of living things is formed by natural forces.

The membranes tend to shape themselves into certain characteristic forms, such as the plastids, mitochondria, endoplasmic reticulum and agranular (Golgi) reticulum. These forms are found universally throughout cells with quite monotonous repetition. The specification of them is essentially a matter of control of the foldings of membranes. The forms do not seem to be structurally very elaborate and it is not beyond the bounds of our imagination to suppose the conditions under which they arose.

All the 'components' of the cells seem to share this quality of a certain basic simplicity. Thus proteins exist with a fibrous, globular or worm-like tertiary structure. The sensitivity of cells to changes around them often depends on certain types of molecules that are capable of making large changes in 'structure' as a result of small energy changes. The great sensitivity of both the nose and the eye depends upon such molecules. In this sense we may be said to 'smell' light or 'see' the chemical structure of a molecule. It is remarkable to find nearly identical pigments in the eyes of animals such as octopuses and men, whose common ancestor lived more than 500 million years ago. Moreover the pigment is exposed to the light by spreading it out on membranes, though with some differences in geometrical construction.

To give one more example, the motile hairs that are attached to cells of all groups of animals from the earliest protozoa to the vertebrates all have at their base a ring of nine fibrils, with two at the centre of it. We do not understand the significance of this, but it is obviously a plan that was adopted very early and has persisted ever since. Once the specification of it had occurred, it was repeated over and over again.

These hints of basic simplicities of organisation are a welcome sign that this most obscure phase of the origin of life may be beginning to yield to study. But it would be a great mistake to imply that we can do more than glimpse the processes by which the most elaborate, continually acting chains of reactions can have been put together to produce even the simplest 'eobiont' ('dawn life') from the probiotic slime. We can say, however, that it is

likely that its composition was 'regulated' in some way by the action of nucleic acid chains, themselves self-replicating and able to organise the production of proteins. The 'structure' of the eobiontic material may have been based from an early stage on membranes. In association with such molecules as chlorophyll the earliest metabolic cycles would have begun. Acts of synthesis proceeded with the energy obtainable from the sun by virtue of the complicated sequences of reactions within the web of membranes.

Although the picture must remain vague, it is just possible to glimpse how an organised, steady-state, self-maintaining macromolecular system may have arisen. Once this property of persistence or homeostasis appeared, it proved to be most resistant to destruction in spite of cataclysmic changes in world conditions. While the land mass has split up and its component continents have drifted apart and been raised and lowered many times, life has gone on with rather little basic change in its composition. Cell structure is much more stable than mountain structure. Is this in spite of the fact of continual turnover or because of it?

THE LIVING UNIT WHICH GROWS AND DIES

Superimposed upon the basic stability and similarity of composition of life there has of course been the continual change that we associate with reproduction, ageing and death and the associated phenomenon of change of type or evolution. In the discussion so far we have not considered the problem of the division of living matter into 'individual' portions, nor that of the duration of existence of each of these. It seems likely that from a very early stage the living matter would have been broken up into small droplets, allowing interchange with the materials around. Each such 'individual' would presumably 'grow' and eventually divide into two. This property or tendency to increase seems to be one of the basic features of the molecular condition that we call the living state. The cycles of activity that allow the building-up of the enormous macromolecules do not proceed to a culmination and then stop. The sequences have been so selected that they continue, or, putting it the other way round, out of all the combinations that have occurred certain have proved to be self-continuing. The tendency not merely to persist but to increase seems to be a fundamental condition of the self-maintenance. Life and growth are inseparable.

If the individual portions of living matter are to grow they must ultimately divide, and we reach the conception of 'individual' living organisms, which lies so deep that we cannot easily imagine 'life' apart from 'living things'. Probably from very early stages such individual masses have indeed been fundamental units. Elaborate metabolic cycles continued by virtue of the organisation achieved through 'control' by the nucleic acids of the operations of a limited volume of matter. These concepts of the individual and its control are certainly among the most fundamental for biology, and indeed for all thinking. All of our concepts of direction and purpose in evolution may be said to derive originally from these actions of the nucleic acids.

These molecules so regulate the sequence of synthesis within each mass of living matter that it 'develops' through a series of stages, each able to continue life under some particular set of circumstances. These actions thus constitute what we recognise as a pattern of 'instructions'. It is by virtue of them that the system is homeostatic, self-maintaining within the environment in which it finds itself. The instructions indeed provide for actions to be performed to meet changes imposed by the environment, at least within a limited range. For example, there are provisions for healing and repair of parts that become damaged or lost.

THE SURVIVAL OF THE FIT

The fact that the instructions provided by the nucleus ensure survival is to be explained according to the classical theory of natural selection. The principle that Wallace and Darwin suggested is commonly called the Survival of the Fittest, with the conception of a struggle for existence. The essential point, however, is the Survival of the Fit. The organisms that exist do so because their hereditary systems are those that unfold a pattern of action that ensures continuity of life in a particular environment. This has been achieved by the selection of just those combinations of nucleotides that produce proteins which, acting as enzymes, ensure the right syntheses at the right times. Any combinations that fail to do this become eliminated.

Once again we see the importance of selection from a set of

possibilities as the principle by which life is possible. In this case we can say that selection in the past of certain nucleotide combinations provides 'information' about the actions by which life was made possible in the past. With this information the organism makes forecasts as it develops. The tadpole hatching from the eggs placed by the frog in the water emerges with a tail for swimming and a mouth and intestine suitable for utilising plants. Later its genes instruct that its thyroid gland shall so enlarge that the processes of metamorphosis are initiated, the tail is absorbed, legs grow and the little frog comes ashore, forecasting that it will find there flies to eat.

The 'selection' that makes this fitness possible must, of course, be made among a large number of alternatives. This is perhaps the basic reason for the universal association of life with growth and reproduction. The delicate balances among the cycles that ensure the continuation of the macromolecular systems are achieved by keeping the 'information' continually up to date. This is done in higher animals by the process of sexual reproduction. Every individual carries the hereditary instructions in the form of the various arrangements of the bases in nucleic acids strung out along the chromosomes of the nuclei of its cells. For example the particular arrangement along a certain stretch of one chromosome ensures that the production of the enzyme that makes black pigment in the iris is delayed. The result is a child with blue eyes. The method by which such 'genes' are inherited was, of course, first revealed by Mendel and is now quite thoroughly understood. The chromosomes are in pairs (23 pairs in man) and each individual receives one of each sort from each parent. Since every individual has a different set of genes (except for identical twins) this combination of parental influences provides a continual supply of slightly different individuals in the population. It is by selection among these that the hereditary information is kept up to date.

There are also other means for increasing the variety in a population. Since each parent contributes only one of the two chromosomes of each pair to each child, it is clear that when the germ cells (eggs and sperms) are formed, the members of each pair must separate. This is provided for by a process known as meiosis, halving. But before they separate, the chromosomes coil around each other and by breaking and reunion actually interchange material. Clearly this operation serves further to mix up the genes. We do not inherit whole chromosomes from our grandparents. Each chromosome that we receive is a combination of the two that our father or mother received from their parents.

The new combinations provided by this crossing over and sexual reproduction ensure that the genes present in a population are continually reassorted to make fresh types of individual. In addition the genes themselves may change, by the process known as mutation. The materials that compose the nucleic acids are, of course, subject to chemical change like any others. In particular, as a result of bombardment of radiation of high energy, small changes may occur in the molecules. This is the means by which X-rays and radiation from an atomic explosion produce their effects. By slightly altering the arrangement of the bases they cause the nucleic acids to produce different proteins – in other words, when development proceeds in the next generation the new forecast is made that a particular protein combination will survive. Since the nucleic acid is self-replicating, the mutant will be a permanent feature of all surviving descendants. Its effects may not, in fact, appear in all descendants because of the phenomenon of dominance. The two genes of each sort in any one individual quite often do not have equal effects; one, the dominant, may completely mask the other, which is said to be recessive. It is important, therefore, to distinguish between the genetic constitution of an individual, known as the genotype, and the outward appearance or phenotype.

MEETING THE DEMANDS OF LIFE

Sexual reproduction is then the chief means by which it is ensured that sufficient new types appear to allow the 'adaptation' of the race to changes in the conditions around. Of course we have to distinguish between such hereditary adaptation and the changes that the individual organism undergoes in response to new conditions. Muscles grow more powerful and bones stronger with use. When an individual goes to high altitudes the amount of haemoglobin in his blood is increased to give extra oxygen-carrying power. There are hundreds of such mechanisms of adjustment in the body and each of them must itself depend upon a hereditary mechanism. The haemoglobin is increased at low oxygen pressures because there is an enzymatic mechanism ready to provide the

synthesis when needed. In general these mechanisms all depend upon the principle that the absence of some substance is the key that unlocks the system that ensures its synthesis. Obviously the organism cannot make *anything* to order, but it is surprising how even bacteria, if starved of one particular amino-acid and provided with only tiny amounts of the raw materials necessary, will produce up to one-third of their own weight of the enzyme needed to make the best of a bad job.

We can thus see that the living system adjusts itself to ensure survival by actions with various time scales. The changes that we ordinarily call 'physiological' mostly happen in a few seconds or minutes. Our hearts beat faster when we run, or the kidneys excrete water when we have taken too much. All the hundreds of such changes are also 'adaptations', just as much as the increase of haemoglobin at high altitudes. The only difference of principle is that the latter change takes several days. Similarly the re-sorting of the hereditary material also produces adaptational change, but with a still slower time scale. As we walk along there is a very precise mechanism in the ear that predicts the probable course of events a few seconds ahead and keeps us upright. To predict what form of organism will be required a hundred or more years hence is much more difficult. For this we have the mechanisms of reassortment of the genes, selection and the consequent change that we call evolution. It is important to recognise that there is no difference in principle between adjustments on all these different time scales. All of them take place by means of special mechanisms controlled by the DNA of the chromosomes. The 'mechanism' of meiosis operates under these instructions just as do all other living processes. Its 'function' is to ensure persistence into the future. Evolution is part of physiology; it is adaptation occurring on a long time scale.

We should not perhaps exaggerate the extent of the 'forecasting' that is involved in evolutionary change. Looking back at the geological record we see the persistence of organic forms through immense periods of time, and their slow change. But the system at any one time is not really able to forecast what conditions will be like 10,000 years later. It is true that some zoologists speak of a characteristic as providing a 'pre-adaptation' or 'prospective adaptation'. Thus fishes acquired air bladders, open to the exterior, which were useful as lungs when the animals came on to the land. But the bladders were of course also useful to the fishes in the water, for respiration, flotation or both. A true 'pre-adaptation' would presumably be useless in the present condition of the species, but might be useful in the future.

In a sense all mutations have this character, since they occur at random and not as a direct 'adaptive' response to an environmental change. In any case no population elaborately prepares itself for conditions thousands of years hence. It continually adapts itself to the conditions *now*. Evolution is going on every second of the day. Paradoxically in this sense it is an adjustment on a shorter time scale even than changes of respiration and circulation.

EVOLUTION IN ACTION

One of the main sources of knowledge about evolution has been the geological record. It was indeed the extinct mammals of South America that first strongly attracted Darwin's attention to the problem. Ever since that time there has been a tendency to see evolution in its longest time scales of hundreds of millions of years. Study over these long periods is inevitably difficult and often inexact and has left the impression that knowledge about evolution is somehow in a different category from knowledge about, say, respiration.

Fortunately in recent years, by concentrating on less dramatic examples of evolutionary change, it has been possible to show the process actually at work, adjusting the organisms to small changes in their surroundings by small alterations, made over a few generations. The long-term 'forecasting' that was spoken of above is therefore not quite so forward-looking as has been implied. Meiosis and sexual reproduction are not some almost magical mechanism foretelling the future 10,000 years hence. They are a mechanism which ensures the presence of organisms able to live next year.

In fact evolution is going on all the time, just as is respiration. It would be unfair to imply that Darwin himself showed a failure of imagination in this respect. He was equally impressed with medium- and even short-term changes in animal form. An example of the former is the finches of the Galápagos Islands, which he visited. These birds have been studied again recently by Lack. The islands were formed by volcanic action, probably not more than

two million years ago, the nearest mainland – Ecuador – being 600 miles away. They contain a remarkably limited fauna, apparently all descended from a few species that have managed to cross from South America. There is one species each of rat, bat, snake, lizard and gecko. The tortoises, after which the islands are named (from the Spanish *galápago*, a tortoise), have grown to become animals of giant size, presumably favoured by the absence of other large vertebrates. But the most remarkable animals of all are the birds. They are all descended from seven mainland species. Several of these have become differentiated into distinct species on the various islands. The finches have become so differentiated that they are classified into fourteen species and four genera. Many of these species have habits quite unlike those of normal finches.

The Galápagos Islands therefore give us an example of evolution producing adaptation to various habitats over a period of some two million years. This is a short time for geologists, but still immensely long in relation to individual animal lives. It is interesting that in recent years several examples have shown the process of evolutionary change at work over much shorter periods. In several industrial countries there are species of moths that have become black during the past hundred years. The Peppered Moth (*Biston betularia*) sits on the bark of trees with wings spread. It has a mottled colour that breaks up its outline. In 1850 the first black ('melanic') form was found in Manchester, and since then the condition has spread so widely that in the north of England nearly all Peppered Moths are black.

When the trees are blackened the moths gain an obvious protection from predators. But the situation is not quite as simple as that. The experience of those who breed moths is that the melanic forms are hardier in the sense of being less liable to disease. It seems that under normal unpolluted conditions the black forms cannot spread in spite of this advantage, but they readily do so when the trees are black.

Even shorter-term evolutionary changes have been observed. Cain and Sheppard investigated the fact that small snails (*Cepaea nemoralis*) show great variety of colour and banding. By marking snails and placing them on various types of background they showed that thrushes remove a greater proportion of yellow shells if the background is brown but of pink or brown shells on a green background. The percentage of yellow shells found at a thrushes' 'anvil' actually got less as the background became more green between April and June.

This is an especially interesting situation because it explains the advantage for persistence that is conferred by the presence of several varieties in the population (polymorphism). It is advantageous for the snail species to have brown members available in the winter and spring, yellow ones in the summer. Such polymorphism enables the species not only to remain alive at different times of year but also to exist in various habitats. Thus the *Cepaea* in beech woods included few yellow ones, those on open downland many.

THE ADVANTAGE OF BEING DIFFERENT

Human populations are polymorphic for many factors. Thus it has been known for some time that certain characteristics of the blood ('blood groups') are controlled by pairs of genes. These genes occur in different proportions in different parts of the world and anthropologists have used this fact as a means of identifying 'races'. The assumption has been that there is no 'advantage' in belonging to one group rather than the other. E. B. Ford challenged this assumption in 1942, saying that this 'balanced polymorphism' must mean that each of the genes conferred some as yet unknown advantage. Since then Aird and his colleagues have shown that those who carry the A gene of the ABO series more frequently suffer from cancer of the stomach than those who lack it.

Many other examples of such differential susceptibilities have been suspected and are under investigation. An especially interesting one is the blood condition known as sickle-cell anaemia. This is produced by a gene that is fatal if handed down by both parents. Nevertheless it exists in a considerable proportion of certain populations because those who get it from only one parent have a greater resistance to subtertian malaria than do individuals who lack the gene altogether. Where malaria is not a problem the gene has already become less frequent, for instance among those of African descent in the United States.

Many things can be learned from this example. For instance, the gene is common among Negroes, the Veddoids of southern India and Greeks; genetic similarity is therefore not necessarily an indication of 'racial affinity'. An even more important thought is that the best way of removing a 'dangerous gene', such as that

for 'sickling', is not by such 'eugenic' methods as differential mating, but by removing the environmental hazard (in this case malaria) that allows the gene to remain in the population (Kalmus).

However, the general point from these human examples is that they show that adjustment of gene frequencies is proceeding continually in the population *as a result of selection of those who have some advantage*. Curiously enough we know more about the population genetics of man than of any other animal. For such work man is the species of choice because medical needs provide sufficient incentive to collect the vast mass of data that is necessary. It is often suggested that human evolution has come to a halt. In fact we have more evidence that evolution is continuing in man than we have for most species.

It may be objected that the cases mentioned are very special ones, concerning only some detailed features of human pathology. This attitude shows a failure of imagination of the immense complexity of the factors that are involved in homeostatic adaptation. Creatures remain alive only because of the infinitely detailed adjustment of thousands of conditions, each of them individually no more and no less 'petty' than those that allow a child to survive with malaria or make an adult develop cancer.

If species are continually adjusting to the environment in this way, it follows that it is likely to be very difficult to follow the detailed course of evolution in the fossil record. This is demonstrated by the curious fact that the more we know of this record the more confused it becomes. The classic example of evolution, that of the horse, seems fairly clear because it has been reconstructed from a relatively small number of specimens. Where all, or many, individuals leave traces the palaeontologist is confused by an immense array of fossils. It is very difficult for him to tell which have evolved from which, and he must distinguish many parallel lines. This is particularly unfortunate because series of fossils are especially useful for such practical purposes as the identification of levels at which oil or natural gas may be found. 'Oil geologists' usually owe their expertise to knowledge of fossils. Unfortunately the detailed sequences are inextricably confused when a considerable proportion of all animals that die leave their shells. When gradual changes can be detected, it usually remains impossible to specify their adaptive significance.

MORE COMPLEX, MORE EFFICIENT

Although there is no doubt of the fact of evolutionary change, it is less easy to be sure in identifying what we may call its general feature or direction. Each population becomes adapted to continue existence in some particular 'niche'. Does the sum of such changes produce any consistent trend? The question obviously becomes insistent as we approach the point of considering our own position as creatures who are part of this living whole.

One measurable change has been an increase in complexity. A man has more parts than an amoeba, and an amoeba more than a virus. The best measure of such complexity is perhaps expressed by the number of 'words' in the instruction books by which the organisms operate. A moderately large virus has a DNA chain of 200,000 bases, that is about 60,000 'words' each of three letters and specifying one amino-acid on a protein chain. A bacterium has up to 20 million base pairs, 100 times more than the virus, and a man 1000 times more again, perhaps 6000 million 'words' per cell.

It is easy to make these comparisons for organisms that are very different, but much harder when they are near together. Is a man a more complex organism than a horse? Perhaps even here we might make some analysis on a basis of the number of alternative actions taken in unit time, and hence in the amount of information passing through the system, though this would be difficult to evaluate.

It is not of course simply the number of parts or of actions that is significant but the fact that higher organisms are able by virtue of their complexity to survive under conditions in which the simpler ones cannot live. Each organism only maintains itself distinct from its environment by active processes that prevent its dissolution. For example, any organism living in fresh water has to pump out the excess water that continually tends to flood into it. The more remote an organism is from the conditions in the sea, the more work is needed to keep it intact. For example, a terrestrial animal is a watery system in a nearly dry environment. It must do work in order to obtain water.

The complexity of higher animals is due to the fact that they have more devices than the lower ones enabling them to live under more 'difficult' conditions. The genetic 'instruction' book of higher

organisms is longer than that of lower ones because it contains the information necessary to construct these various parts. In particular, higher organisms contain more devices than the lower for obtaining information direct from the environment by means of receptors and storing some of it in brains.

As evolution has proceeded, life has by these means almost certainly expanded into areas where it was not possible before. Probably in so doing the total amount of living matter has been greatly increased, though I know of no attempts to estimate this 'biomass' at different stages of evolution. The total energy turnover required to maintain this increased mass is also probably greater, both absolutely and relatively to the mass of living matter involved. Living under more 'difficult' conditions requires a greater energy expenditure, just as it requires more equipment and more energy to move 100 yards up Mount Everest than to walk 100 yards along the street. Notice that we do not say that the climber is 'less efficient', because we judge that the task he is attempting is more difficult.

This colonisation of conditions ever more remote from those in which life first arose is probably a consequence of the mechanism for ensuring evolutionary adaptation. If there is continual pressure of excess growth, reproduction and variation, then some types suitable to an increasing range of conditions will appear. It does not follow that *all* lines will show this tendency, and some regression is indeed observed. But the overall change in the direction of increased complexity is unmistakable. V. Foerster has expressed it neatly by means of estimating the total number of cells within each type of organism as a function of the number of cells within each individual. There are more cells within a few dozen human beings than there are among all the Protozoa in the world. There is therefore evidently an advantage to be obtained from increasing complexity over and above the fact that it makes life possible under conditions not colonised before. This may be the explanation of the fact that in general whole types or grades of organisations tend to flower and then become almost or quite extinct – replaced by their own descendants. Thus the reptiles replaced the amphibians nearly completely and were in turn largely replaced by the birds and the mammals. We get the impression that there is a continual increase in 'efficiency' within each environment, as well as a continual venturing into new environments. Thus a modern fish is probably more agile in the water than was its heavily armed palaeozoic ancestor.

However, it is notorious that in some organisms there has been a secondary simplification of structure during evolution. This is especially true of parasites. The bag-like structure of *Sacculina* attached to the abdomen of a crab is a creature that once was a barnacle, reasonably highly organised with legs and shell. Even in such cases we must be careful not to overlook complexities that may be present. *Sacculina* has a system of roots spreading throughout the tissues of the crab. Moreover the species is distributed by a larva that retains the typical organs of a crustacean. The genotype must therefore retain most of its old complexity and add new features to make feeding upon the host possible. The life of parasites is not necessarily an easy one; it requires its own special skills.

Perhaps a more important question is whether the course of evolutionary change proceeds at a constant rate or at rates that show any recognisable pattern. Clearly there cannot be a constant rate for all, because some species remain almost unchanged, perhaps for hundreds of millions of years. This perhaps also is to be expected if we look upon evolutionary change as the making of adjustments to meet small changes in the world around. If the environment remains the same, so do the animals. The difficulty is to see what features of the environment of lungfish and coelacanth have allowed them to remain unchanged, whereas other fishes have altered so greatly. They do not seem to be especially protected.

In any case it is not a question of explaining the existence of those few outstanding relict forms. They are only the extreme examples of a large range. It is often maintained that various types of evolutionary change can be recognised. Large 'evolutionary breakthroughs', such as the colonisation of the land by amphibians in the Carboniferous period, may indeed have involved great and rapid changes. After such periods of explosive evolution it is held that the rate of change decreases and some few forms may linger on for a very long time. Simpson, Westoll and others have devised various means of measuring evolutionary rates in order to test such hypotheses and these do indeed seem to show the effect. It still remains difficult, however, to be sure to what extent we are confused by the difference in time scales, for instance between individual lives and those of species. We are apt to speak of

evolution as 'rapid' because, say, ten populations that are recognised as genera appeared in the course of about 20 million years. The only meaning that should be attached to 'rapid' here is that the number of genera is greater than appeared in the preceding or succeeding intervals. It can have only a very attenuated reference to the extent of change between one generation and the next or even over periods such as 100,000 years, which are already immensely long in relation to individual lives. Retaining a sense of proportion is difficult in such matters and becomes especially important in dealing with the history of human evolution.

Knowledge of man's ancestry is made unusually difficult by the fact that since Primates are mostly arboreal they are seldom preserved as fossils. It is estimated that out of some 100 extinct genera of Primates only ten are reasonably well known. However, several types of early Primate have survived with relatively little change from Eocene times, 70 million years ago. By study of them and of such fossils as are available, it has been possible to reconstruct the general outlines of primate and human evolution, though there is still considerable controversy about many phases.

THE EVOLUTION OF MAN

It has often been observed that in respect to many features man and all other Primates are rather primitive mammals. This is especially striking for those who study limbs, since there are few other mammalian orders in which five digits remain in both hand and foot. The most specialised forms of movement among Primates are perhaps the mechanisms for swinging with the arms ('brachiation') found in gibbons and some monkeys. The foot of man is also a remarkable structure, suited to carry us over rough and smooth ground and large and small obstacles, Similarly the teeth of Primates do not show the extreme specialisations found in herbivorous animals for grinding hard plants or in carnivorous ones for cutting flesh. The teeth of some of the more primitive Prosimians retain the triangular form that is found in many of the early mammals and was probably associated with an insectivorous diet. Indeed the oriental tree-shrews (*Tupaia*) are so exactly intermediate between other Insectivora and Primates that there is still much dispute as to which group they should be put in.

The tree-shrews show us that the taking to the trees was certainly one of the most significant steps in primate evolution. The earliest mammals probably had a long snout, like modern shrews; they may have been nocturnal and with poor eyes. With the adoption of life in the trees, smell became of less and sight of greater importance and this has been the tendency for Primates ever since. Moving among the trees they need to detect events far away with their eyes and learn to interpret them correctly. The acquisition of information in this way is perhaps even more important to an animal in the open, exposed to the full light of day, than for one semi-nocturnal or burrowing.

Moreover movement requires to be even faster and more precise among the branches than upon the ground. Development of the brain has certainly been one of the most striking features of Primates, quite apart from man.

Already by the Miocene period, 25–10 million years ago, there was quite a large fauna of well-developed monkeys (*Mesopithecus*) and apes (*Proconsul*). The monkeys already possessed a characteristic pair of ridges on their molar teeth, such as are still found in monkeys today. The teeth of apes and of man show no such ridges, but a set of four or five low cusps. We can therefore be reasonably sure that our own line of evolution diverged from that of monkeys at least 25 million years ago. Some people believe that this divergence was very much earlier. The late Professor Wood Jones was so much impressed with the similarities between man and the strange Prosimian, the oriental tarsier, that he considered this to represent an early stage of the human line, diverging 60 or more million years ago.

We can probably assume that there was a common monkey–ape–man stock in the Oligocene period, and that the monkeys separated in the early Miocene. The time of separation of apes and men is much more problematical. Some of the Miocene apes have been supposed to be especially likely candidates for human ancestry. *Proconsul* was not a full brachiator but there is no real evidence that it walked erect. In fact throughout the whole Miocene and the 10 million years of the succeeding Pliocene period there is no good evidence of the existence of any creatures that could be considered human. The absence of evidence is suggestive, since fossils of other mammals are of course common and even those of apes are not very rare. However, it must be emphasised that the chances of fossilisation of a Primate are so small that

absence of fossils cannot be considered to prove the absence of living types.

At the very end of the Pliocene or beginning of the Pleistocene period, one to two million years ago, there are, however, two groups of fossils of creatures that are man-like, but definitely not men of the modern type. The Australopithecines, found by Dart, Broom and Leakey in Central and South Africa, were creatures that ran or walked on two legs, but had brains only one-third the size of man. Moreover their teeth, though in pattern like man's, were large and set in a massive jaw. Some of the skulls had heavy, projecting brow ridges.

These fossils are of quite extraordinary interest, especially since the bones are found associated with primitive tools, almost certainly made by them. Whether we can decide to call these creatures men is a matter of definition; a more interesting question is whether any of them actually contributed genes to our modern human stock. There is at present no definite answer to this, but specimens from Java and Peking dated (rather arbitrarily) at about 550,000 years ago are more definitely human. The brain-size was now two-thirds that of modern man but the shape of the head was very different from ours. The forehead was very low, suggesting slight development of the frontal lobes, but the occipital lobes were large. The pelvis and femur show that these people walked nearly or quite erect. Their teeth and jaws were still very massive.

Throughout the greater part of the time since the Peking and Java men lived there is evidence that people even more similar to ourselves existed. Some of the skulls still show 'primitive' features (heavy teeth, etc.) and are placed in a distinct species of the genus *Homo*. There are several sorts of these Neanderthal men, some of them quite like the typical *Homo sapiens*. Their brains were perhaps a little larger than those of modern man, and they walked fully erect.

THE LEAP TO MODERN MAN

The work of archaeologists made it clear many years ago that through most of the Pleistocene period men were hunters, living nomadic lives in small family groups. Probably this condition persisted for nearly a million years. It was only some 10,000 years ago that man ceased to be a food gatherer and settled agricultural and urban life began. Unfortunately there is very little evidence to show what precipitated this fundamental change. In his evolution since settlement began man has shown astonishing powers, and it would be of great importance to discover how and when these manifestations began.

The particular features of human life are interwoven with each other and it may be that they all evolved together. Social co-operation between individuals is an essential for all typical human activities. There seems to be no reliable evidence of human life in groups larger than the family earlier than 10,000 years ago. The basis for group co-operation was presumably some glandular and cerebral changes that repress the more extreme manifestations of conflict. The large development of the frontal lobes in *Homo sapiens* suggests the possible basis for this restraint. Experimental and clinical studies have shown that these lobes serve in part to maintain the balance between inquiry and restraint that is appropriate to creatures that acquire their information by learning from co-operating with each other. The earliest *sapiens*-type skulls, such as that found at Swanscombe near London, already showed high frontal lobes more than 250,000 years ago, but no evidence of complex social life is found until hundreds of thousands of years later.

The delay in the onset of maturity is another characteristic of man of whose origin we know nothing. Bolk suggested that many human features (including the large frontal lobes) could be explained by assuming that man is an ape who never grows up. It is indeed an ingenious thought that we are human because we fail to grow to full apehood. However, even if true it does not tell us when, where or why the change occurred.

Although these physical changes were perhaps necessary for the emergence of man as we know him, it is possible that the critical change was the alteration in the information transmission system. The large brain, social life and long childhood made it possible for each individual to learn from many others. Through speech began a system of what we may call multiparental inheritance, each individual learning from many. With writing, man produced his unique contribution to the evolutionary process – an extracorporeal information store. Squirrels and many other animals lay up stores of food, but only man has invented non-living codes, by selection

from which he can store his information outside his body. One of the strangest features is that in the various languages there should be so many *different* manifestations of this unique capacity.

Most of the features of human life that we have been considering find some parallels elsewhere in the animal world. The social life of bees and ants is notoriously developed and has made them very successful. They also have elaborate communication systems. Moreover we have perhaps been rather slow to recognise the high degree of co-operation between individual mammals and the complexity of their social and communication systems. There is probably still much to be learned about these in, for example, whales and in other Primates.

But by the making of a coded record of experience, for transmission to any others who can learn the code, man has begun a wholly new phase of evolution.

Note: The original chapter includes fine illustrations and captions, for which the student is referred to the book, cited on p. xi.

THE PSYCHOLOGY OF THE EMERGING SELF

Everett D. Erb and Douglas Hooker

THE STUDY OF HUMAN BEHAVIOUR

In thinking about human beings and why they act as they do, everyone tends to have a built-in do-it-yourself kit for explaining behaviour. From the historical and cultural influences of the past and from his own unique mosaic of experiences, each person develops perception of customs, of social institutions, of skills, of duties, of what is meaningful and important in life, and of what is natural, right and proper in the behaviour of man. According to Muller[1] Herodotus once stated that 'if one were to offer men to choose out of all the customs of the world such as seemed to them the best, they would examine the whole number, and end by preferring their own'. The same preferences are likely to be true for explanations of human behaviour, and everyone, from the highly trained psychologist to the beginning student, is faced with the difficulty of handling this problem.

You would be quite correct in resenting any implication that the ideas presented in the following [pages] are automatically correct and will necessarily have to replace your personal concepts. Instead, the ideas that follow, although considered to be important by some of the people who have made a profession of studying human behaviour, are written with the assumption that you will examine and assess them carefully, put them to use in life situations on a tentative and experimental basis, and perhaps augment them with your present views whenever and wherever experience would justify doing so.

In order to be able to incorporate into your value system the concepts that will be presented, it will be necessary to approach your reading with a relatively open mind. You need to be willing to review what you know, consider carefully how you came to believe it, and examine what evidence exists for supporting your present position. Where differences occur between your ideas and those in the book, you must be willing also to decide for or against either position on the basis of the available evidence. Disagreement may confuse and even distress at first, but the awareness of diversity, complexity and fluidity, and the ability to accept uncertainty, are among the basic marks of the serious student and the scholar. Unanimity is comforting but it is also often spurious and misleading.

Strangely enough, children, who from two to five years of age approach life with an inexhaustible curiosity, often grow up to be adults who never ask the question, 'why?' In fact, sometimes even college students seem to regard everything they are asked to think about as an imposition. What has happened to their enthusiasm for learning? It may well be that on the road to adulthood they were offered too many pat answers which they were required to accept and not challenge, and they soon learned that life was simpler when they did not raise embarrassing questions. The students' spirit of inquiry may have been damaged also more or less permanently by a thinly disguised fear of questions on the part of their parents and teachers. As Friedenberg has stated,[2] 'Adults who do not basically like and respect adolescents – and this includes a large proportion of those who make a career of working with them – are badly frightened by the increasingly democratic relationships between adolescents and adults that are coming to prevail in our society.' Moreover, perhaps both parents and

teachers have talked down to young people and to newcomers in a particular area of study on the pretence that they could not understand true complexity and, consequently, must be fed a kind of intellectual pabulum. Is it so strange, then, that some students react passively to such a diet? Probably most people would assume that very young children would not be able to learn several languages simultaneously because the demand on the children's intellect would be too great. It may be, however, that what children can comprehend and learn has been seriously underestimated. One of the authors lived in Port-au-Prince, Haiti, for a number of years and was amazed at the number of pre-school-age children he encountered who spoke English, Creole, French and sometimes Spanish or German. It would seem to be a reasonable hypothesis, then, that children or adults can learn much more than they actually learn or are allowed to learn. The crucial difference seems to be a matter of motivation. The children in Haiti learned the languages because they wanted to in order to communicate with members of their families, friends, neighbours and servants, and not to satisfy the expectations of parents and teachers or to reach some vague and far-off future goal.

METHODS OF EVALUATING KNOWLEDGE

The use of common sense

For many people, common sense appears to be one of the great arbitrators concerning the correctness of ideas about human beings and how they function. The concept needs to be carefully re-examined, however, before being too readily accepted as the ultimate criterion for behaviour. The first word, *common*, implies that the knowledge is shared by all. Everyone, it is assumed, partakes in the common belief. The second word, *sense*, is related to the way in which people are presumed to have come to share in this commonalty, i.e. through seeing, hearing, tasting, smelling, touching and so forth. In actuality, however, this shared sensory experience known as common sense has frequently been proved to be incorrect. At one time, for example, common sense assured man that the world was flat. As he looked across a stretch of level ground it was difficult for him not to have the experience of flatness. Despite this obvious reaction, some individuals of uncommon sense kept insisting the world was round. Further evidence, above and beyond the first and more obvious sensory impressions, substantiated the latter hypothesis. The exploits of John H. Glenn, Jr, and the other astronauts confirm what today is accepted by all people, and what was once considered to be a matter of common sense concerning the shape of the world has now been changed.

Another example of the fallibility of common sense may be seen by asking you to guess at the answer to the following problem. Imagine you know a person of well-above-average intelligence who has been blind since birth. He is twenty-one years of age and has completed the undergraduate work in college. Among the many things he has learned is the ability to discriminate between a triangle and a square by using his sense of touch. Through tracing the outline of these figures he can discriminate one from the other effortlessly and without error so that the concepts of squareness and triangularity are thoroughly familiar to him. Suppose, furthermore, that he undergoes an operation which results in vision being given to him for the first time in his life. His visual ability in the physical sense is suddenly normal. How long do you expect it would take this individual to discriminate *visually* between a triangle and a square? Fifteen minutes? Two hours? Two days? Common sense suggests that for most people the correct answer would be found within these limits; but how long does it actually take? According to Hebb,[3] Senden's evidence indicated that the shortest time in which normal discrimination can occur is about one month.

It is quite possible that many of the ideas that you are sure you know about human behaviour may be nothing more than a tenuous expression of common sense, but if you are to seek the truth, then the tendency towards relying on the irrational certainty of common sense needs to be replaced with a healthy doubt and too easy and pat answers by a questioning spirit. 'May it not be the greater merit of questions that they lead not to answers but to new questions, and the new questions to others, and they to others yet?'[4] Not having the answers may seem to be fatal, but failure to search for meaningful questions will almost certainly prove to be disastrous.

The use of authority

A second method for evaluating knowledge is through the use of authority. Frequently, an idea is considered to be true or not true on the basis of the position taken by a famous man or authority.

Because a person is an expert in one specialised field of study, however, it does not mean that he is qualified to make recommendations in all other areas of knowledge. Moreover, further understanding and information may cause the initial position of an authority to become inappropriate and out of date. For example, a few years ago medical experts considered it to be good practice to bleed patients in order to cure a variety of ailments, and even barbers industriously implemented this knowledge. Today obviously such methods would be considered incorrect and the practice of medicine is never entrusted to people without the necessary qualifications. You are cautioned, therefore, against too quickly and blindly accepting ideas, purely because they are presented by recognised authorities.

The use of empirical evidence

Still another, and more reliable, method for ascertaining truth is that of obtaining empirical or observable evidence. The story is told of Aristotle, sometimes credited with the first use of the empirical method, that he settled the argument over how many teeth a woman has by the simple experiment of asking his wife to open her mouth while he counted bicuspids and molars. There were pitfalls on this road to truth, for as the story goes, Aristotle's wife was the exception to the rule and possessed either more or less teeth than the average woman. Whether true or false, the story illustrates the point that truth is never easy to discover. Sometimes it depends on the point of view of the person seeking it, sometimes on the method used to obtain it.

Russell pointed out that 'every advance in civilisation has been denounced as unnatural' and he offered the following guidelines to the searcher after truth. First, he felt that if a matter can be settled by observation, it is important that the observation be made and the situation not be assumed to be true. Second, he warned against passionate convictions. If another person's position causes anger, the excessive emotional expression is probably an indication of bias and uncertainty. In fact, the greater the anger, the more likely that there is some unconscious doubt about the validity of the person's position. Third, Russell stressed the importance of becoming acquainted with the opinions and ideas that are held in social and intellectual circles outside the ones in which a person lives, and fourth, he cautioned against opinions that flatter one's self-esteem[5]. It may be concluded that intellectual integrity consists of deciding difficult questions on the basis of as much evidence as can be found and on leaving the question open-ended and undecided when the evidence is not conclusive.

In an attempt to gain a broad and inclusive perspective and to reach into areas of study outside the fields of psychology and sociology, illustrations can be drawn from the ideas of historians, philosophers, artists, social critics and literary figures. Thomas Mann, for example, opened his novel *Joseph and His Brothers* with the words, 'Very deep is the well of the past'. In a very real way, an awareness of the past enables a person to remove himself from the immediacy of his own concerns and to see his hopes and beliefs in the light of all men's hopes and beliefs. Consideration of Mann's words brings to mind the possibility that what each person thinks he knows is just as fragile and fleeting as what men, five hundred years ago, were sure they knew. The words 'Orville, that thing will never fly' linger hauntingly as a reminder of the danger of premature certainty. On many occasions in the past, people have been wrong about man and his potential capabilities, but nevertheless he still must be seen in a broader perspective than his immediate life situation.

The passage which follows, taken from an adaptation of a series of articles written by Russell Lord and published in the 1941 issues of *The Land*, invites us to view an imaginary movie describing the historical and geological developments on earth for the past 750 million years. Starting at midnight of New Year's Eve and running continuously, the film will last for exactly one year. Thus every second of the movie represents 24 historical years and the average person's life will flash by in three seconds.

For the first few months of the film, the viewer sees only a series of rather monotonous changes and transformations in the land masses. During April, however, single-celled organisms begin to appear and, by the latter part of the month, some of them become multicellular. In May the appearance of the first vertebrates, although still in the form of aquatic life, is observed. By June the large resources of limestone, gas and oil available today are in the process of being formed under the shallow seas that cover over half the land area that presently constitutes North America. There is still no sign of vegetation, however, and it is not until the middle of July that the first land plants become noticeable.

In August the most obvious feature of the film is the presence of an abundance of fish in the seas and waters. During the latter part of September the dinosaurs, which will become dominant for the next couple of months, begin to appear. One of the highlights of the film occurs in October when the mammals start to arrive on the scene. In November the dinosaurs disappear, and by December the mammals have replaced them as the dominant influence on animal life. The land is now covered with trees, plants and grasses, and erosion is limited to specific areas. Most of the rivers are crystal clear and on Christmas Day a stream flowing to the south-west begins the erosion of what is known today as the Grand Canyon of the Colorado.

Man has still not made his appearance as the picture moves into the remaining few days in December. The spectators become somewhat concerned over the possibility that for some reason he may have been forgotten. Finally, around noon of the last day in December, a large, awkward creature – Pithecanthropus, the Java ape-man – makes his appearance. He lives in a hazardous manner, menaced both by excessive changes in weather conditions and the wild animals that surround him. Until eleven o'clock in the evening he remains the only type of human being. Then he is succeeded by the Neanderthal man who is followed shortly afterwards by the Cro-Magnon man and then by the Neolithic man.

Only five or six minutes before the end of the film, civilisation makes its belated appearance. The Egyptians, Babylonians, Hebrews, Chinese, Greeks and Romans follow each other in rapid succession. At one minute and seventeen seconds before midnight, Jesus is born. He encourages men to love one another, and for the next minute his followers increase in large numbers. There is much bloodshed among these people, however, concerning matters of belief and doctrine.

With less than thirty seconds to go, Columbus discovers the New World while the Declaration of Independence is signed only a few seconds before the conclusion of the film. The scene is changing rapidly. The virgin forests are destroyed by man and by fire, and the soil, man's most valuable possession, is being washed away to the oceans. At the same time man is seen accumulating wealth in the form of pieces of printed paper that represent a relatively useless metal that he keeps hidden away in vaults and other safe places.

During the last seven seconds, the number of human beings increases at an amazing speed. Their food supply decreases: more than half of them are hungry and some are starving. They are involved in a continuous series of crippling wars which causes them to annihilate one another. Some of the people are hostile, destructive and criminal. Others are sick and dying. Nevertheless there are other persons who have healthy and constructive effect upon human existence. During the final few seconds of the picture, they are seen to be engaged in a vital struggle against the destructive influences that are gradually gaining control. Until man came on the scene, the constructive aspects of nature were beginning to win out over the destructive forces. The deleterious behaviour of man, however, turned the tide and caused the destructive powers to gain ground again. As the picture reaches its last three seconds, the outcome is still uncertain. It is during these three seconds that you and I were born.

Against such a background of geological time, man seems insignificant. As Russell states, he 'is a brief episode in the life of a small planet in a little corner of the universe, and that, for aught we know, other parts of the cosmos may contain beings as superior to ourselves as we are to jellyfish'[5]. Perhaps in the light of such a perspective, it becomes easier to accept man's knowledge as still evolving and tenuous and to appreciate the opportunities which exist for further development and understanding.

PRESENTATION OF GOAL-DIRECTED BEHAVIOUR

One of the assumptions that is made about behaviour is that it may be understood most effectively when one knows the goals a person sets for his life. Fig. 1.1, although a rather complex schematisation, attempts to give some tentative order to the various factors which influence the selection of goals.

At the left side of the figure is the human being with his inherited characteristics and individual potentialities. Each person is seen as a self-activating entity, responsible for creating a part of his own environment and capable of becoming not merely a passive recipient of external stimuli, but an individual who is able to express himself uniquely and differently from other people. This ability to create a part of one's environment may be seen even in the behaviour of new-born babies in a maternity ward,

THE PERSON
Individual differences
Internal and external determinants

INTERACTING WITH HIS PERCEIVED ENVIRONMENT

ATTEMPTS TO ACHIEVE A

HIERARCHY OF GOALS

I. Orientational Goals
II. Physiological Goals
III. Love, Belongingness and Psychological Safety Goals
IV. Achievement Goals
  Respect
  Wealth
  Power
  Skill
  Enlightenment
V. Subjective Satisfaction in Living
  Well-Being – Level I
    Psychotic, neurotic and over-conformity
  Well-Being – Level II
    Integration and actualisation

WITH VARYING DEGREES OF ENERGY MOBILISATION

LEARNING POTENTIAL
  Differential reinforcement
  Generalisation
  Discrimination

EMOTIONS
  Positive, ambivalent and negative feelings

ADJUSTMENT MECHANISMS
  Goal substitution
  Reality distortion
  Goal withdrawal
  Self-punishment

THROUGH AVAILABLE

INSTITUTIONS
Man seeking values through institutions using resources

AND INTERPERSONAL RELATIONSHIPS

Expressive
Manipulative
  coercion
  coaxing
  evaluating
  masking
  postponing
Altruistic

Fig. 1.1 A schematic presentation of goal-directed behaviour

for it is possible to distinguish differences in the quality, the intensity and the duration of their rage, in the strength of their sucking response and, in short, in a bewildering array of behaviour patterns which exist long before substantial learning from the environment has had an opportunity to take place. At such an early age, the infant is already beginning to act in a way which is characteristically his own and consequently he is creating, at least partially, the particular environment from which he will learn and to which he will be expected to react.

The environment surrounding a person is represented by the external determinants or the powerful cultural and social forces to which the individual will be exposed during the course of his existence. He grows up learning to speak a particular language or languages as opposed to others and he experiences the company of certain individuals and groups instead of others. For example, a father obviously influences greatly his child's learning experiences inasmuch as the father's income, vocation, religion, politics and so forth all have impact upon the child. In the early and formative years, at least, these factors are beyond the infant's control and he may experience them only passively. Such influences are so numerous and so strong that there is a tendency sometimes to ignore the other factors which accompany them and contribute to the shaping of behaviour. For instance, the father also reacts emotionally to the child and his behaviour is influenced in part by what the child does or does not do. If the infant cries continuously and keeps the family awake at night, he may bring out feelings of hostility in the father, whereas another child, who cries less, is more likely to enjoy a warmer relationship with his parent. The

quality of the relationship, in turn, affects what is learned and valued by the child so that his attitudes become interrelated with the type of experience he knows with his father. Where the relationship is warm and loving, the child will not need to react strongly against the values of the parent. Where the relationship is characterised by coldness and friction, there is a much greater chance that the child will want to express some dissatisfaction and disapproval.

Similarly, in every human relationship there is an intermingling of the influence of the environment with the behaviour patterns which the person himself is responsible for developing. As the experience of the person increases and he gains a deeper understanding of his particular needs, he begins to make choices for himself and to determine different goals for his life which will be important to him with varying degrees of intensity. To the extent that he is successful, he is able partly to establish his own environment and to reduce the danger of his becoming a slave or victim of the particular setting in which he finds himself.

The specific goals which a person chooses may be seen as being arranged in a hierarchy in which the basic and more primary ones must be achieved successfully before new and higher goals may be realised.[6] Thus, in a growing and developing person, continuous new goal emergence becomes essential while, in contrast, excessive concentration upon any one goal may be seen as an indication of at least partial failure on the part of the person to cope with his environmental experiences.

The classification of the goals which are important to human beings may begin with the most basic one of orientation. Most Americans, for instance, are oriented towards the responsibilities of good citizenship, and consequently they were appalled by the behaviour that took place in the Los Angeles riots of 1965. No matter how much they might desire to accumulate material possessions, it would be impossible for them to do so by stealing and looting. The people involved in the riot, however, did not share in this orientation, as evidenced by the remarks of one woman in response to having her picture taken in the process of looting, 'It ain't your store. What do you care?' Perhaps it is possible to understand, therefore, that depending upon their orientations, people will behave differently in the same situations and also the same way in different situations. A classic example of the influence of differences in orientation has been provided by Snygg and Combs:[7]

> Several years ago one of the authors was driving a car at dusk along a western road. A globular mass about two feet in diameter suddenly appeared directly in the path of the car. A passenger in the front seat screamed and grasped the wheel, attempting to steer the car around the object. The driver tightened his grip and drove directly into it.
>
> In each case the behaviour of the individual was determined by his own phenomenal field. The passenger, an Easterner, saw the object in the highway as a boulder and fought desperately to steer the car around it. The driver, a native of the vicinity, saw it as a tumbleweed and devoted his efforts to keeping his passenger from overturning the car.

Since all behaviour is determined at least in part by the person's perception of the stimuli around him, and not wholly by a simple automatic response to such stimuli, the orientational goals are considered to be the most important and basic ones in the hierarchy.

The next goals are related to the visceral, sexual, cognitive, sensory and motor processes, all of which are included under the category of physiological goals. For the most part, these goals are characterised by an urgency that demands gratification before the person is able to move towards the achievement of other goals.

Following the physiological goals is the need for love, belongingness and psychological safety. Love is expressed most effectively through behaviour which allows the individuals involved in a relationship to achieve the particular goals that are important to each of them. Feelings of being psychologically safe are developed when love in this sense is supported by an orientation which is congruent with the objective reality of the outside world.

Once the person's more basic needs have been satisfied, he is able to turn, in his own unique fashion, towards attaining the goals of respect, wealth, power, skill and enlightenment. Provided his behaviour is oriented intrinsically or, in other words, is seen as having merit in itself, the person may be expected to enjoy a sense of fulfilment over his accomplishments. On the other hand, if the attainment of these goals is seen only as being valuable because it

makes further achievement possible, then the person experiences little satisfaction and his behaviour becomes just another incident in an endless series of failures to accomplish something that is personally meaningful and significant.

Finally, there is the general goal of well-being which may be seen at two different levels. The first level includes a range of people with mental health varying from the seriously disturbed person or psychotic at the lower end of the scale to people who are reasonably happy, content and successful at the upper level. As measured by most personality tests and as seen by the man in the street, the people in the latter group seem to enjoy a fair amount of good mental health, and they frequently make important contributions to society.

There are also some people, however, who seem to enjoy a more creative approach to life, and a second level of well-being is postulated to account for the behaviour of those persons who possess a degree of maturity which is achieved by only a few members of society. The behaviour of such people is characterised by the knowledge that what they can be, they must be.

Furthermore, a person approaches his goals with varying degrees of intensity. A simple and perhaps not too accurate analogy may by made with an automobile which requires a steering wheel, equivalent to direction or goals, and also gasoline, which is seen as intensity or emotion. If high-octane aviation fuel is used in the automobile, the engine will quickly burn up and cease to operate effectively. On the other hand, if a cheap or inferior brand of gasoline is used, the motor will knock and also perform inefficiently. In one case the gasoline is too strong and consequently disruptive to the reaching of a destination or goal, while in the other case the gasoline is too weak for effective movement to take place. What is needed, of course, is a brand of gasoline which is neither too strong nor too weak, but one of a more optimum strength that will permit the automobile to operate effectively. Similarly, a person needs an intensity of emotion which is also neither too strong nor too weak but, instead, is adequate enough to make it possible for him to learn and to utilise his potentiality in a way which will allow him to reach his particular goal or destination.

When behaviour does not lead to the achievement of goals, certain devices, usually referred to as adjustment or defence mechanisms, are used to offset the feeling of failure. Sometimes the adjustment mechanisms are both healthy and realistic in that they bring about the formation and subsequent achievement of new and more appropriate goals. On other occasions, however, they have the opposite effect of causing the person to ignore or misinterpret the reality of his own experience, thus providing support for the negative connotation that is frequently associated with the term defence mechanism. In an ideal world in which all people are capable of achieving their goals and therefore do not encounter failure, a negative view of adjustment mechanisms would be appropriate. In the world as it actually exists, however, in which the conditions of life almost guarantee some degree of failure in goal achievement, the use of adjustment mechanisms in helping a person overcome his feeling of disappointment after failure may be of considerable value. While excessive dependence on adjustment mechanisms may inhibit goal achievement, not utilising them at all may lead also to disappointment and frustration. What is needed again, of course, is a moderate or optimum use that will help the person in eventually obtaining the goals that are important to him.

The goals that a person seeks, the intensity with which he desires them, and the various adjustment mechanisms he uses while pursuing them may be seen as existing at different levels on a scale of awareness or consciousness. Some of the goals a person desires to achieve, for example, are very clearly known to him and may be considered as being at the upper level of the scale. Other goals, which are much more vague and difficult to recognise, may be placed at the lower end of the scale, for they are below the level of consciousness in what is generally referred to as the unconscious. Sometimes the goals are at a higher level than the intensity and adjustment mechanisms, while at other times the opposite is the case. In fact, any conceivable combination or arrangement may be realised. However, the more a person is consciously aware of his goals, knows clearly how important they are to him, and understands the adjustment mechanisms he uses, the more he is able to make his own decisions, hold particular beliefs and behave in a rational manner. Conversely, the more these factors are at the lower and unconscious level of the scale, so that the person is vague and uncertain of them, the more he will have difficulty making his own decisions and formulating consistent attitudes

and the more he will be prone to function irrationally instead of rationally.

In order to achieve the goals that are seen as being important, a person must relate to other people through the various institutions that are available in society. Some relationships are primarily manipulative in that other people are coerced or encouraged into behaving in ways that will lead to desired goals. On other occasions, when a person's actions are not specifically directed towards the attainment of a goal, the behaviour is seen as being predominantly expressive in nature. For still other people, needs are satisfied mainly through helping other people reach the goals which are important to them. Such behaviour is altruistic and is frequently found, for example, in counsellors and psychotherapists. It is very closely related also to the goals that emerge in Well-Being Level II.

The study of human behaviour would be easy if it could be classified precisely in terms of the foregoing discussion and the outline in Fig. 1.1. Unfortunately, this situation is not the case. Instead, Fig. 1.1 is included with the hope that it will become a tool or guide by which you may examine your own behaviour and that of the people around you.

There may be a feeling, at this point, that human behaviour is much too dreadfully complicated. Consequently, perhaps it is advisable to offer some words of reassurance. Many of the terms which seem vague will [later become more clear]. Furthermore, if you have learned that there is no one simple approach or answer, then undoubtedly you have taken the first step along the road towards sophistication in the study of human behaviour.

Notes

1. H. J. Muller, *The Loom of History* (New York, 1958).
2. E. Z. Friedenberg, *The Vanishing Adolescent* (Boston, 1960).
3. D. O. Hebb, *The Organization of Behavior* (New York, 1949).
4. J. Ciardi, 'Manner of Speaking', *Sat. Rev.*, vol. 45: 9, no. 22, 1962.
5. R. E. Egner and L. E. Denonn (eds), *The Basic Writings of Bertrand Russell* (New York, 1961).
6. Editor's italics.
7. D. Snygg and A. W. Combs, *Individual Behaviour* (New York, 1949).

The excerpts which follow are taken from Alexander d'Entrèves's book 'The Notion of the State'. You will see that he deals with many problems which are directly related both to our main question 'Why do men live in society?' and to the supplementary question 'What do men gain by living under government?'; he discusses such issues as the relationship between 'law' and 'order', the question of 'legitimacy', and the concepts of 'negative' and 'positive' freedoms.

These extracts are, of course, only a part of Professor d'Entrèves's argument. The book as a whole attempts to present in a new perspective what still is the basic notion of political theory, that of the State. Its main contention is that, in dealing with this notion, three different levels of approach ought to be distinguished: the level of might (or force), the level of power and the level of authority. Using this distinction, the author discusses some of the most significant versions of politics which have been put forward in the past and are held at the present day by scientists, lawyers and philosophers. The State, in his view, is not merely a system of force as conceived by political realists. As the rightful holder of power, it is essentially a legal construction. But legal theory in turn is unable to provide an adequate account of authority. Political theory is once more called upon to consider, and possibly to solve, the problem of political obligation.

LAW AND ORDER

Alexander d'Entrèves

THE first, the easiest and the most widespread legitimation of power is that which involves the need for order, and describes order as the greatest benefit which can be secured by the State. We need only turn to current language to find the word 'order' used in this sense. 'The preservation of order', 'the forces of order', the 'new order' as opposed to the 'old': how often we come across these and other similar expressions in newspapers and speeches, not only in those of a markedly conservative bent! Order is a magic and exquisitely political word, loaded with an evocative power which the word 'legality' does not possess. No doubt, in English at any rate, law and order are usually linked together. But it is the notion of order, not that of law, which arouses emotion, and never fails to awake a response in people's imagination and desires. For indeed, in the popular mind, order is opposed to disorder, or, as is also said, to anarchy. Order is normality, safety, peace – even peace at any cost. Thus understood, order does not too obviously coincide with legality. The general who reported to the Czar 'Order rules in Warsaw' is not likely to have been a lawyer. To him, law simply meant martial law. Yet it was in the name of law that he justified his recourse to force. A certain degree of ingenuity, perhaps even a kind of professional distortion, is needed to enable us to declare that order is in its own way a rough, primitive form of legality, and to see in it at least the beginning of that control and normalisation of force which we singled out as the distinctive mark of the State. The appeal to order, and to the benefits it entails, implies the transference to the plane of values of the notion of power, a notion which, as we have seen, is in itself indifferent and neutral.

The classic example of such a transfer is still that of Hobbes. As we saw, Hobbes identified the State with legality, that is with an order which guarantees safety and peace. But this order is at the same time a value: indeed, it is the highest value attainable in this world by men. It therefore constitutes not only the essence of the State, but its justification. We need not, however, be Hobbesians – that is we do not need to accept Hobbes's premisses, his predominantly pessimistic view of human nature – in order to take the step from maintaining that the State exists to guarantee law and order to demanding that this order should be established because it is a good thing. There is a sense in which every political philosopher, one could almost say every good citizen, indeed every man who willingly obeys the laws, implicitly makes Hobbes's reasoning his own. He does so in so far as he 'accepts' order not only as a fact but as a value, that is, as something which not only exists, but should rightly exist, and from which he derives a positive advantage.

Thus, if we turn to political theories of the past, we find hardly any which cast doubt on the merits of law and order. The myth of the 'state of nature' was neither a subversive nor an anarchical myth, a means of disproving the value of order. It certainly was not such for Hobbes, who used it to stress the need for the State. But neither was it such for 'those who of old sang of the age of gold and of its happy state', taking a more optimistic view of human nature. Men in that condition of happy innocence obeyed the law without constraint and lived together in peace and harmony. But, like the hierarchy of Heaven, they bore witness to an order, to that order which, according to St Augustine, is the condition of peace: *pax omnium rerum tranquillitas ordinis* ('Peace for all things is the tranquillity of order'). Nor are utopian dreams necessarily anarchical. Utopias, the models of a perfect society, in which at certain moments in history men sought an escape and refuge from reality, do not imply a denial of order; far from it. They are expressions of an aspiration after a better or different order, after an order which may well be more rigorous and exacting than that which actually exists and is laid down in positive legal systems.

But, it may be objected, it is one thing to dream of the lost happiness of a fabulous age, or to imagine perfect constitutions in far-off, unknown lands, and quite another to justify the State as a coercive machine. In the one case we assume the existence of an order which rests on the evidence of reason, in the other we refer to an order which is distinguished by the use of force. But this objection clearly misses the point. For the value of order is something quite different from, and independent of, the way in which order takes shape and is established. There will always be a question of deciding whether order is or is not something desirable before we decide whether it can and ought to be imposed by force or achieved by some other means. And here everyone seems to agree, Hobbesians and anti-Hobbesians, political realists and seekers after Utopia: that order is desirable and necessary; that it has a positive value; that it is what distinguishes men from beasts and makes their life worth living.

We have to travel to the very threshold of our modern world to find this view deliberately challenged. What we then come up against is a depreciation of order, or to be more exact an emphasis on its purely negative side. Order sets limits to the power of the individual to act: hence it entails the sacrifice of something that is unique and precious. It is contrary to the 'true nature' of man, and as such it is an evil: what is worse, an unnecessary and perhaps an avoidable evil. Such are the views which crop up for the first time in the eighteenth century, almost in contrast, as it were, to the final establishment of the modern State, to its complete monopoly of power. They are set forth in fables and myths according to the taste of the century. Yet these fables have not lost a certain poetical truth, an endearing freshness and candour. They deserve to be remembered not only because they foreshadow later theories, whether anarchical or liberal, deliberately hostile to the State, but because they register certain demands with regard to order and law which have become part and parcel of our attitude towards power. Three authors in particular come to mind in this connection, each telling a different tale, but each pointing to similar conclusions: that order is not necessarily a good thing, that men might be better off without it and that there are certainly a number of values which legislation cannot secure and can even hinder.

Of these three authors, Montesquieu, with his Story of the Troglodytes, is the least remote from tradition. The Troglodytes, a small people of Arabia, having killed their king decide to devote their lives solely to the pursuit of their particular interests. The consequence is complete confusion and insecurity – Hobbes's state of nature. Even the doctor, who has been refused fair payment, refuses in his turn to cure them when an epidemic overtakes and decimates them. The few survivors are thus converted to 'humanity', to 'justice' and to 'virtue'. Happiness and prosperity, peace at home and bravery in face of the enemy, flourish again. But the Troglodytes decide to give themselves a leader once more. They turn to a venerable old man, who weeps on learning that the choice has fallen on him:

> 'I shall die with grief for having seen the Troglodytes born free, and now to see them in subjection. . . . I see, O you Troglodytes, what it is. Your virtue begins to be too heavy for you. In the condition you at this time are, you must be virtuous in spite of yourselves. You cannot subsist without it, but would fall into the miseries of your forefathers. This yoke seems hard to you. You had rather submit to a prince, and obey his laws, less rigorous than your morals. You know you may then satisfy your ambition, acquire riches, and languish in vile luxury; and

provided you avoid committing great crimes, you will have no need of virtue.' He stopped a moment, and his tears flowed faster than ever. 'Ah! what would you have me do? How can I lay my commands on a Troglodyte? Would you that he should do a virtuous action because I command it? Would he not do it himself, and from an instinct of nature only?'

The moral of Montesquieu's story seems fairly simple. Government is redundant for men who are truly virtuous. In fact, government may be an excuse for dispensing with virtue altogether. If men followed reason there would be no need for the State. This moral is not substantially different from that of the old fable of the golden age. For what is redundant is the coercive machinery of the State, but not order as such; quite the contrary. It is the rule of reason that makes the peaceful, happy coexistence of human beings possible. They are better off without the State, but only in so far as they are reasonable.

To find the very benefits of law and order denied we must turn to another myth, that of the happy savage. We must turn to Rousseau's startling interpretation of the state of nature, which is radically different from that of all his predecessors. For Rousseau the nature of man does not lead, as it does for Hobbes and Spinoza, to the war of all against all. But neither does it entail a rudimentary, peaceful society, as it does for most theorists of the state of nature. Man left to himself does not stand 'in need of his fellow-creatures' nor has he 'any desire to hurt them'. A simple hut and the few tools needed to sustain daily life are all that is necessary for leading a 'free, healthy, honest and happy' existence. Self-interest, the multiplication of needs, contact with other men, have laid the roots of slavery. In a word, it is society which has depraved and perverted man. History is but the tale of the progress of inequality.

But this, as is well known, is not for Rousseau the end of the story. The *Discourse on the Origin of Inequality* finds its counterpart in the *Social Contract*. Rousseau's final message is not the destruction of order, but the building of a new and better one. The basic problem of politics is that of finding the means of enabling man to be 'as free' in society as he was in the state of nature, of helping him to recover the equivalent of his natural independence under the form of civil liberty. This is the problem which Rousseau faces in the *Social Contract*, where the citizen is guaranteed against any form of 'personal' dependence by his 'total alienation . . . to the whole community', i.e. to the State. Thus, we shall see, Rousseau ended by definitely turning his back on the happy savage, and perhaps even opening the road to the ethical State or to modern totalitarianism.

But Rousseau's peculiar interpretation of the state of nature does not for that reason lose its importance nor its fascination. The myth of the happy savage struck deep into the imagination of his sophisticated contemporaries. The gospel of the return to nature could easily be turned into an open invitation to anarchy. An appeal of this kind can be found, for example, in Diderot's *Supplément au Voyage de Bougainville*, which gives perfect literary expression to the longing for escape, to the quest for the exotic, of an age which, under a veneer of civilised refinement, was pervaded by thrills of rebellion and by an unquenchable thirst for renewal. We almost seem to hear already the chidings of a Marx or the oracular utterances of a Pareto. 'Look carefully at all political, civil and religious institutions. If I am not deceived, you will find the human race crushed from age to age beneath the yoke which a handful of rogues has sought to lay upon it. Beware of the man who would impose order. To create order always means to make oneself master of others at their expense.'

According to Diderot, among the nations of Europe only the Calabrians, the inhabitants of the southern tip of Italy, had managed not to let themselves be imposed on by 'the flattery of legislators'. He praised their 'anarchy', the genuine and happy condition of primitive, as yet uncivilised, peoples. He pointed out that 'their barbarism' might well be 'less vicious than our urbanity', and that, at any rate, 'the heinousness of a few big crimes about which so much fuss is made' was amply compensated by our 'petty wickedness'. Significantly enough, a similar argument was used, only a few years after Diderot, by one of the earliest prophets of Italian nationalism, Vittorio Alfieri. Steeped in the reading of Machiavelli, Alfieri openly undertook to defend the 'enormous and sublime misdeeds' which were Italy's special privilege. They were to him a sign that Italy 'possessed even now, more than any other European country, an abundant supply of fierce and fiery spirits. Such characters needed only the room and the means to achieve great things.' Here, indeed, an 'ethics of heroism' is proclaimed,

which was soon to become a cherished motif of the Romantics, and could lead one day to the cult of the superman and to the open defence of violence.

All that remained to be done, in addition to eulogising anarchy, was to pronounce the vindication of the anti-social nature of man, of egoism, of the pursuit of personal profit, with no further regard for the traditional view, which Montesquieu's Troglodytes had to their cost learned to know and respect, that virtue and justice are the foundation of society. This is the theme of the most caustic of all eighteenth-century apologues, Mandeville's *Fable of the Bees*. Mandeville's purpose is to show that 'private vices . . . may be turned into public benefits'. 'Fraud, luxury and pride', combined with 'hunger', are the springs of political life. The prosperity of the State, like that of the hive, is founded on the greed of its members, on the exploitation of one set of human beings by another. 'Such were the blessings of that State; Their Crimes conspir'd to make them Great.'

It is, however, noticeable that Mandeville, in spite of all his cynicism, still uses the traditional vocabulary of morals. Vice remains vice, evil is still evil. It took some time for these vices to be called economic virtues. Mandeville's point is merely that 'what we call evil in this world . . . is the grand principle that makes us sociable creatures'. What he does in effect is to propose a new set of values. He turns the tables on Hobbes's state of nature. Man's basic instincts no longer need to be thwarted by the iron claw of Leviathan. They are a constructive force, full of the promise of progress. All that is needed is to let them loose, the less hampered the better. Competition, the struggle to win power and wealth, should take the place of the old aspiration after law and order. In a world such as this the State is reduced to a purely negative role. It may well turn out in the end to be nothing but an instrument of power for those who take control of it, and an insurance of survival for those who are unable to do so.

Three myths, three different worlds. In each of them the modern reader can recognise as in a distorting mirror some familiar trait. In each there are to be found particular features which have in some way contributed to the shaping of modern man's attitude to the State. In an age such as ours, which has known and knows the most extreme forms of idolisation of State power, it may seem paradoxical to suggest that the State is devalued. And yet it might well be asked whether the two views which hold the field in State-theory today – political realism, which reduces the State to force, and legal positivism, which acknowledges its presence in every effective system of law – do not in fact imply such a devaluation. The truth is that there is no point in talking of devaluation or valuation in considering theories which deliberately take their stand on the plane of neutrality, in discussing a type of argument whose purpose, as I have repeatedly pointed out, is to remain accurately descriptive. But this does not alter the fact that, as we saw, value judgements are often made even by realists and positivists: Machiavelli and Hobbes are outstanding examples. Nor does it alter the fact that a judgement of this sort is certainly present in that substitution of the idea of order for that of legality, which we observed taking place in everyday language and from which we started out in this chapter.

A value judgement was certainly made by all those – and they are the majority of political philosophers – who in the course of centuries stressed the concept of order as the first and foremost legitimation of power: though not any kind of order, but only one that is 'just' and 'right'. Here the Augustinian definition of order is more significant than any other: 'Order is a disposition of equal and unequal things giving to each its place.' This definition in turn clearly echoes that given by the Roman lawyers, of justice as a 'firm and enduring will to give each one his right'. In both cases the accent is no longer on what is, but on what ought to be. Order is not merely a situation of fact, it is an ideal state of affairs, something to be striven after, which provides a standard for judging a given situation. Order, indeed, is equivalent to justice; and since the State is the guarantor of order, justice alone can be the foundation of the State, the reason for its existence: *iustitia fundamentum regnorum*.

This approach is entirely different from any we have hitherto encountered. From this new standpoint the State is no longer merely a force that imposes itself, nor a power exercised according to any law whatsoever. It is a power 'authorised' to exact obedience in view of the attainment of certain definite ends. Thus the analysis of the concept of order has changed the mood of our whole inquiry. Henceforth we must be prepared to meet definitions and theories which are anything but neutral. A momentous step has been taken in political discourse, the step from description to prescription.

NATURE AND CONVENTION

Alexander d'Entrèves

In what sense and under what conditions is it possible to define as 'right' and 'just' the order which is achieved in the State by means of the legal system? The question, put in such terms, may sound preposterous to modern ears. Yet it is one that has been asked ever since the dawn of civilisation and that still, however differently worded, troubles the mind of man. To provide an answer to that question has long been considered the special task of political philosophers; it is an answer that must be given, were it only to prove that political philosophy still exists. In fact, the issue is twofold. To ask in what sense law and order are said to be 'right' is clearly a question about definition. It implies the admission that, at a certain point in our attempt at defining the State as the guarantor of law and order, we have abandoned the purely factual approach and have introduced some reference to values. But to ask, on the other hand, under what conditions the order set up by the State is 'just' involves us in an assessment of those values. Clearly, we must first ascertain if, and when, and how, value judgements are resorted to in defining the State, before we proceed further to examine the nature and character of such judgements.

To define the State as a 'right order' may mean one of two things. It may mean either that the State is the bearer of a value called 'justice', or that it is simply an instrument, a means of achieving that value. At any rate, in both cases State and justice are inseparably linked. Aristotle's statement on this point can be taken as paradigmatic. 'Justice,' he says, 'belongs to the *polis*; for justice, which is the determination of what is just, is an ordering of the political association.'

Let us briefly follow Aristotle's argument. At the beginning of the *Politics* we are told that only in the State can man achieve the perfection inherent in his nature. Outside the State man is a being either above or below humanity, 'either a beast or a god'. Man is by nature a political animal (*politikon zōŏn*): this means that political life is the natural condition of mankind, whereas its absence is an impossible, almost a monstrous, condition, just as a man who is unable, or has no need, to share in the benefits of political association can hardly be called a man. For 'man, when perfected, is the best of animals, but if he be isolated from law and justice he is the worst of all'.

Everything obviously hinges on the concept of nature. For Aristotle it is the end which determines the nature of beings. The nature of a thing is its condition in the final, perfect, stage of its development. To say that the State is natural does not, therefore, merely mean that in fact men do live in political societies. Aristotle certainly does not refuse to resort to factual confirmation in order to prove the 'naturalness' of the State. Experience teaches us, he says, that man is the most sociable of all animals, the only one gifted with speech; that men are differently endowed, and that the household, with all its inequalities, is the kernel from which historically the State draws its origin. But the chronological order is reversed when we come to consider the end, that is the justification, of the State. Even though the naturalness of political life is empirically warranted, and even though it is possible to retrace the growth of the State in the development of human associations, the historical perspective is completely altered by means of the concept of nature described above, which puts the centre of gravity at the point of arrival, not at the point of departure. Therefore, whatever may have been the sequence in time, the State is logically and morally prior to, indeed more 'natural' than, the family or the village out of which it has grown; for it is the condition of the family's and the village's existence and well-being, just as the whole is the condition of the existence and well-being of the parts. The State is the bearer of a good which has the value of an end, for the individual as well as for all subordinate groups. This good is justice. 'Justice', Aristotle repeats in the *Nicomachean Ethics*, 'exists only as between men whose relations to one another are governed by a system of law.'

Yet, in presenting the State as the achievement of justice, Aristotle had no intention of justifying any and every kind of State. He had too deep a sense of the variety and relativeness of concrete political forms, and of their remoteness from the ideal type or model which alone would ensure the complete coincidence of justice and the State. Such a coincidence does not and cannot occur except in the 'good' State, just as the virtue of the ordinary citizen is not, and can in no way be, the same as that of the good man that, is as virtue in general, except in a good constitution. The

fact remains, however, that justice is the essential value-connotation of the State, and that the State is not only an instrument of this value, but its embodiment. It is a value that is, so to speak, immanent in the State, and can be attained by the individual only through a full and active participation in political life. The 'ethical' nature of the State is the corollary of this immanence: the idea that the State not only guarantees an external, formal order, such as the one which is set up by law, but provides the conditions for a life lived in accordance with virtue, i.e. for the realisation of moral life. Thus Aristotle assigns the highest possible task to the State, at the cost of belittling morality: for the fulfilment of the moral ideal, tied as it is to the City, is bound to be peculiar and partial, that is, lacking that universal character which we usually associate with moral values as such.

The political philosophy that has for centuries dominated Western thought, far more effectively than Aristotle's, gives an entirely different account of the relationship between State and justice; and it is one that has to a much greater extent affected our history and our ways of life. How did this change come about in the centuries immediately following Aristotle, how were new values revealed, and by whom? This is not the place to retell the highly dramatic story, in what would necessarily be a totally inadequate way. What matters is to emphasise that the discovery of these new values led to a new and quite different way of justifying the State. While in Aristotle the attribution of value to the State had found expression in the doctrine of its 'naturalness', now the State no longer appears as something natural, but as something conventional. The argument follows a line expounded by the Sophists and taken up again by the Stoics; it is based on the contrast between *physis* and *nomos*, between what is natural and permanent and what is conventional and subject to change. Nature no longer means a final point of development, but on the contrary an initial condition; and it is to that condition that we must turn if we want to understand why the State came into being, the reasons that justify its existence, and the role it plays in human affairs. Thus, in defining the State as a conventional institution, the value-connotation was no longer in-built, but brought in as it were from outside. It was a matter of knowing the particular grounds which made the State necessary, or useful, or desirable, and which alone could account for the change from the original to the present conditions of mankind. In the traditional version of the theory the State exists for the implementation of certain values which are universal and unchanging – 'natural' in the sense that they correspond to the very nature of man. These values find expression in the 'law of nature', that law of which God himself, according to Cicero's famous definition, is the author, and reason the interpreter, which is not bounded by the walls of the City but is valid for all nations and all times.

At first sight this may well appear to be a wholly negative attitude, one of almost deliberate depreciation of the State. Deprived of the value of being itself an end, the State is reduced to being simply a means of attaining values which are given, not made. Justice is no longer immanent in law and order; it transcends the level of politics: it is realised only if and when the State is modelled on a higher order and on a higher law. But a closer examination reveals that depreciation is not a correct description of this kind of approach. Rather, it implies a positive appreciation of the possibilities offered to men of building up a 'just' order, and of the responsibilities laid upon them for doing so. The 'true city', to use a wonderful image of Dante's, may not be achieved here and now; yet it may be possible to perceive, at least in the distance, its towers and spires. Since the State is a human contrivance, the central problem is a problem of origins. Only by assessing the origins of political bondage can man know why he must choose to live in the State and obey its laws.

So numerous over the centuries have been the interpretations of these origins – offered in the various accounts of the State's 'conventionality' – that it would take volumes to list them. In the Christian Middle Ages the idea of sin could offer a perfectly adequate explanation of the transition from the state of innocence to the actual conditions of mankind in political life. It could even colour man's need to submit to law and order with the light of resignation, transforming into a duty what in purely factual terms was a painful necessity and an inescapable fate. With the coming of a new age, the claim for individual independence was on the contrary to lead political philosophers to assert the need for an act of will – a 'social contract' – to explain and to justify the rise of political institutions, as well as to determine their limits and ends. It is only in our own day that the theory of the conventionality of political life appears to have been finally contradicted by the

evidence of the so-called social sciences, and particularly by anthropology. They all seem to agree in maintaining that the existence of some form, however rudimentary, of political organisation is a phenomenon which can be empirically proved, for it is found among all peoples in all ages. Therefore, so the argument goes, there can be no doubt that we ought to speak of the State as something natural, not as something conventional or artificial.

But this kind of argument would have had little hold on traditional political philosophers. And indeed it misses the most important point in the contrast between convention and nature, as well as its real contribution to the modern theory of the State. No doubt the relation between the 'state of nature' and the State (or 'political' or 'civil state') could be and was conceived as a relation of sequence in time. This, at any rate, was the way in which the Stoics and the Fathers of the Church conceived of it; so also did the Roman and Byzantine jurists, as did the medieval political philosophers, who, before Aquinas rediscovered Aristotle, could make no use of the Aristotelian notion of the political nature of man. This way of thinking was further encouraged by the story of Eden in the Bible and by the never-forgotten myth of the Golden Age. Even at a later stage, at the time when the doctrine of the social contrast carried the day, political theorists (many of them, though not all) continued to think of the state of nature as the initial condition, corresponding to the happy infancy of the world, a condition which they believed could be found once more in the newly discovered lands full of virgin promise: 'In the beginning all the world was America', according to Locke's stirring phrase.

But, as had been the case with Aristotle, so with the theorists of the state of nature or of the social contract, the empirical or historical confirmation of the condition referred to as 'natural' was only of comparative importance. What mattered was the value attributed to those conditions as setting up a standard or model, as helping to bridge the gap between what was and what ought to be. As we know, the historical sequence is completely reversed when Aristotle speaks of the State as natural. Similarly, when the Stoics, the Fathers, the medieval philosophers, the theorists of the social contract, or even the modern Utilitarians, take the opposite view, and maintain that the State is not a natural but a conventional institution, what they have in mind is not so much the historical sequence as the importance of showing that the State is a means of achieving certain ends. Hence it matters little whether the state of nature ever existed or the social contract ever took place. A judgement of value could never have been derived merely from the reference to a factual state of affairs. State of nature and social contract were normative concepts: their whole importance lies precisely in this. For only by conceiving the State as a human contrivance could the thought of harnessing Leviathan make any sense. Political institutions could then be reduced to their true proportions, their limits secured, their duties defined. The path could be kept open for the pursuance of values higher than politics, the ultimate values, which in order to be truly universal must apply to all men *qua* men.

The proof that this is not possible when the State is regarded as an end in itself and as the bearer of the highest values emerges clearly from the modern theory of the 'ethical State'. I have mentioned this theory before. Now is the time to come more closely to grips with it, even though, for the reasons I have already mentioned, it may well be doubted whether this theory is at bottom anything but a version of political realism writ large. For, indeed, if according to this theory the justification of the State ultimately lies in the very fact of its existence, then all those points hold good in regard to it which we discussed in considering the glorification of force as the ultimate answer to the problem of politics.

This in fact is the final outcome of the teaching of the man who may be regarded as the greatest exponent of the modern theory of the ethical State, as the real master of all those who in our time have personified the State as the supreme embodiment of justice. At first sight Hegel's teaching seems to follow closely in Aristotle's footsteps. Even his language differs only slightly from that of Aristotle. Both consider the State not only as the instrument, but as the condition, of man's perfection. 'Only as part of the State does the individual possess objectivity, truth and morality.' But, unlike Aristotle, Hegel and his followers do not admit the possibility of a difference between the real and the ideal, between what is and what ought to be. 'On some principle or other, any State may be shown to be bad, this or that defect may be found in it; and yet, at any rate if one of the mature States of our epoch is in question, it has in it the moments essential to the existence of the State. . . . The affirmative, life, subsists despite the defects, and it

is this affirmative factor which is our theme here.' For this reason the State – the modern national State – was for Hegel not only the crown of our whole history but the highest positive expression of the moral life. The 'right' of the State knows no limit but that set by the force in virtue of which it makes itself felt in the face of other States. As in Hobbes, paradise lies under the shadow of the sword. But Hegel differs from Hobbes in that for him the sword not only guarantees survival and peace, but is the symbol of a new morality, of values which can be realised only in the State and to which the individual is wholly subordinate:

> At one time the opposition between morals and politics, and the demand that the latter should conform to the former, were much canvassed. On this point only a general remark is required here. The welfare of a State has claims to recognition totally different from those of the welfare of the individual. The ethical substance, the State, has its determinate being, i.e. its right, directly embodied in something existent, something not abstract but concrete, and the principle of its conduct and behaviour can only be this concrete existent and not one of the many universal thoughts supposed to be moral commands.

The traditional view of the relation between State and justice is here turned upside-down. A new political philosophy, or perhaps rather a new religion, was here proclaimed, which maintained among its dogmas that men were simply tools for creating the power and greatness of the nation, that States were responsible only at the bar of history, and that war was the health of the world.

The tragic happenings which have convulsed Europe within our lifetime ought to have made an end once and for all of this deadly philosophy. There would be no reason to mention it, were it not that it contains an indication of an historical and psychological situation which is without doubt characteristic of modern politics, or at least was so until not long ago. It is no use trying to cope with this situation simply by going back to the older view that the State is a human contrivance, an instrument, not an end. The days when the State could be conceived as a remedy for sin are long past. But so are those when it was conceived merely as a machine deliberately assembled for the pursuit of definite aims. Today, the State is generally regarded as the expression of a cohesive bond which existed even before the legal organisation of power. Political obligation may no longer be surrounded with a religious halo, as it was in the past. But neither is it the result of a simple calculation of the adequacy of particular means to particular ends, as it appeared to the theorists of the social contract or to the Utilitarians. Rather, it is a complex phenomenon, loaded with emotive and irrational elements, not easily reducible to abstract and clearly arguable terms. As the repository of a common heritage, the State is still to many people the object of blind dedication, which may prompt them to great sacrifices but also to great injustice; witness the slogan 'My country right or wrong'. Indeed, the name of this dedication is love of country, and the bond of nationality is in fact, as Hegel saw, the strongest pillar of the modern State as a nation-state. The idea of country and nation thus accounts for the evocative and sentimental aura which surrounds the modern notion of the State. We must therefore examine on what grounds and in what way the legitimation of power has been sought in such concepts rather than in that of justice, or at any rate conjointly with it.

NEGATIVE LIBERTY

Alexander d'Entrèves

WHAT exactly is meant by saying that the purpose of power is to secure human freedom? How and why has the attainment of this end become one of the most effective principles of legitimation of the modern State? These are big and difficult questions to which a complete answer cannot be given in a few lines. But it is at least possible to map some landmarks in the development of political theory to guide our judgement in a discussion the importance of which hardly needs to be emphasised.

Historical evidence does not indicate clearly whether the rise of the modern State was in any sense necessarily connected with one of the various concepts of liberty which prevail at the present time. For Hobbes the 'liberty of the Commonwealth' consists, as we know, in its independence. Machiavelli too, when speaking of political liberty (*vivere libero*, *stato libero*), seems to have this meaning especially in mind. Liberty was to him primarily the

absence of foreign rule, and only secondarily opposition to tyranny. The rule of his 'new prince' was justified in so far as it succeeded in establishing a strong government, not in restoring a kind of liberty which he considered no longer possible, at any rate in Italy, in view of the prevailing corruption of political life.

Thus the ideology which presided over the rise of the 'new principality' seems very far from being one of a 'liberal' type. Yet it is worth while pausing for a moment's reflection before accepting this judgement as final. It can in fact be asked whether the very demand for a strong State, for a firm foundation of law and order, which is so evident in both Machiavelli and Hobbes, and which alone would suffice to explain the rise and success of absolute monarchy in Europe, was not in itself, notwithstanding every appearance to the contrary, a demand for a liberty of sorts. Take the famous sentence at the beginning of chapter xvii of *The Prince*, one of the most outspoken, and perhaps one of the most terrifying, passages in Machiavelli: 'Caesar Borgia was reputed cruel; nevertheless, the cruelty of his repaired the Romagna, united it, and restored it to allegiance and peace.' Consider what are the benefits, the goods, the attainment of which justifies the merciless action of the new prince in Machiavelli's eyes: union, allegiance, peace. These are values we have already encountered. They are closely similar to those which Hobbes invokes in order to prove the need to forsake the state of nature and to enter the 'civil state'.

Indeed, nothing is more revealing than the list Hobbes draws up of all the good things that are missed in the state of nature, with the clear implication that they can be purchased by putting an end to that state, i.e. by setting up a 'common power', a 'State'.

> In such condition [he writes in one of the best-known and most impressive passages of *Leviathan*], there is no place for industry; because the fruit thereof is uncertain: and consequently no culture of the earth; no navigation, nor use of the commodities that may be imported by sea; no commodious building; no instruments of moving, and removing, such things as require much force; no knowledge of the face of the earth; no account of time; no arts; no letters; no society; and which is worst of all, continual fear, and danger of violent death; and the life of man, solitary, poor, nasty, brutish, and short.

Here, as in a photographic negative, we get a clear picture of what are the values, the benefits which, according to Hobbes, are attained in the State. These values are both material and spiritual: they concern the comforts of life as well as the improvement of the mind. They are what in modern terms we should call 'cultural' values; and cultural values are always, in some way or other, associated with liberty, with the free display of human initiative and energy.

True, neither Machiavelli's nor Hobbes's State has much in common with the modern, 'liberal' State. The 'prince' of the one, the 'sovereign' of the other, know no brakes to their power. If freedom simply means independence, they alone are 'free', because their power is limited only by force and their will is the supreme law. Nevertheless, if the purpose of power is to guarantee peace and security, the all-embracing action of law must come to a halt at a certain point, if only to allow the enjoyment of all those goods which are the fruit of peace and security. The justification of law and order lies precisely here. Thanks to them, those obstacles are removed which would otherwise prevent that enjoyment if peace and security vanished. This is a wholly negative approach to the concept of liberty. 'The liberty of a subject', says Hobbes, 'lieth . . . only in those things which in regulating their actions, the sovereign hath prætermitted.' In other words, citizens are free only in that sphere which is not regulated by law: *silentium legis, libertas civium.* Let us, then, give this liberty its proper name. Let us call it 'negative liberty': but on condition that in this respect we number even Hobbes among 'liberal' authors. For indeed the greatest philosopher of Absolutism seems to take a fairly generous view of what is or should be the liberty of the subject, if it includes 'the liberty to buy, and sell, and otherwise contract with one another; to choose their own abode, their own diet, their own trade of life, and institute their children as they themselves think fit; and the like'.

Thus, when from the origin of power we turn to considering its exercise, negative liberty appears as the first and perhaps the most significant justification of the modern State. Under that heading we can begin by listing both the removal of obstacles and the assurance of a sphere of individual independence, described by Hobbes himself as the task of the State. But evidently that liberty could not be held to be secure or complete until a third condition was

fulfilled, until the line was firmly drawn where power must be brought to a halt, until, in one word, the limits of State action were definitely established. It seems almost paradoxical that here once again it is Hobbes who points the way. He does so where he analyses the conditions of a law's being good – since, as we know, laws can be good or bad, though by definition 'no law can be unjust'. Among those conditions is that of being 'needful', by which Hobbes seems to mean that only such laws are to be approved which do not impose useless restrictions or burdens. 'For the use of laws . . . is not to bind the people from all voluntary actions; but to direct and keep them in such a motion, as not to hurt themselves by their own impetuous desires, rashness or indiscretion; as hedges are set, not to stop travellers, but to keep them in their way.' The analogy between laws and hedges is revealing. 'Good laws' are only those which are indispensable for securing men's peaceful coexistence. They should act as boundaries to that sphere which is reserved for the decisions of individuals alone. Once again we are confronted with the concept of negative liberty; but this time that concept is used as a legal standard, as a measure of the limits which the State should not overstep – though, of course, from Hobbes's standpoint, there is nothing to prevent it from doing so. It is fascinating to find the same analogy taken up by Locke. For laws to act as hedges is to him their greatest merit and their proper task. Indeed, it is precisely by doing so that laws, contrary to Hobbes's opinion, are the condition of liberty, since 'that ill deserves the name of confinement which hedges us in from bogs and precipices. So that, however it may be mistaken, the end of law, is not to abolish or restrain, but to preserve and enlarge freedom. . . . For liberty is to be free from restraint and violence from others; which cannot be where there is no law.'

Locke, of course, was and still is the chief philosopher of negative liberty. As a complete formulation of the liberal conception of the State, the *Second Treatise of Government* has probably been equalled only by another classic, John Stuart Mill's *On Liberty*. Mill's essay discusses in more up-to-date language the problem of the defence of the individual against the pressure of social forces unknown in Locke's day. But both writers, like all others who share the same liberal inspiration, are concerned to secure that sphere of individual independence which Hobbes, and many 'absolutist' writers after him, regarded as a gracious gift from the sovereign. No doubt the principles to which they refer in defining the duties of power and the limits of State action are widely different. Locke turns to the law of nature and of reason, which precedes the State, and to the natural rights of man, which are inalienable and imprescriptible. Mill appeals to the principle of utility, 'utility in the largest sense, grounded on the permanent interests of a man as a progressive being'. Further, in defining the ends to be secured the emphasis is different. To Locke the reason for the State's existence, or to use his own words, the reason why men join together in the 'social compact', is the 'preservation of their lives, liberties and estates', the three basic goods which he sums up in one 'general name, property'. Mill, on the other hand, draws up a long list of liberties which society must respect, but then he too sums them all up in a single sentence: 'the only freedom which deserves the name, is that of pursuing our own good in our own way, so long as we do not attempt to deprive others of theirs, or impede their efforts to obtain it.' What matters is that in both cases the legitimation of power is linked to an essentially negative concept of liberty. The purpose and at the same time the limits of State action are determined by the one and only value which must be established and secured: the free and unhampered development of the individual.

It would therefore seem logical at this point to conclude that the concept of negative liberty and the liberal conception of the State founded upon it are nothing but the product of a particular historical period and must be judged and valued as such. This is precisely what we often hear said when Liberalism is depicted as the typical expression of an age which saw the establishment of a social structure where unfettered individualism appeared to be the highest possible good. There is certainly much truth in this contention. There is no denying that the success of Locke's teaching corresponded historically with the rise of the commercial middle classes and the 'possessive individualism' which they professed, just as Mill's popularity was closely associated with the final triumph of *laissez-faire* and of the bourgeois way of life. But since calling Liberalism a 'bourgeois ideology' often conceals a polemical or even a hostile attitude, it might be as well to face the question in its most general terms, and to ask whether the fact that a given political theory was linked to a particular set of historical circum-

stances is sufficient ground for doubting its merits and its validity in a changed and more complex world.

Before even trying to answer this question properly, there is one remark which can and must be made concerning the 'bourgeois' character of negative liberty and of the liberal conception of the State. If this is a 'bourgeois ideology', it is certainly very different, both in its assumptions and in its implications, from that other and in many ways opposite theory – also described as bourgeois – of the governing élites. Unlike modern 'élitism', classical Liberalism was tied, both logically and historically, to the principle of human equality. It claimed for every individual an equal share in the rights of man. Far from expressing a reactionary, or even a conservative, attitude towards the existing social structure, it was an instrument of progress, indeed even of subversion of that structure, of the privileges and discriminations which that structure entailed.

This, however, is not the only merit of the liberal theory of the State. Its greatest achievement was that of having created its own instruments and institutions, which still constitute the firm foundation of the modern State. These institutions have weathered many storms, showing that they can be adapted to the needs of a changed society. They seem, indeed, to have drawn fresh vigour and inspiration from the very criticisms to which they have been subjected, from the very menace of destruction with which they have been faced in recent years. When examining those institutions from a legal angle I pointed out that they were the result both of historical development and of a deliberate commitment and choice. Now the men who made this choice made it in the clearest possible manner, by means of those solemn 'Declarations' with which it became and remains customary to preface constitutional Charters. They laid down the values they considered essential for the legitimation of power, in the form of a vindication of 'rights'. But those rights were, for the most part, negative liberties. They were nothing else than limits appended to the action of the State. Nor did they spring from the brain of these men, like Minerva armed from the brain of Jove. They had been conceived of long before the rise of bourgeois society, long before even the rise of the modern State.

This, then, is the important point to be kept in mind when discussing the political theory of Liberalism: that the notion of negative liberty was the exclusive product neither of abstract theory nor of social conditions of comparatively recent date. The 'rights of Englishmen', which were to be an inspiration and a model for so many Declarations of Rights in different countries, have a long history behind them: they were 'historic' before becoming 'natural' rights' They were won in age-long struggles and set down in famous documents, of which Magna Carta is only the first in time. The English Declarations of Rights in the seventeenth century have no abstract or universal character. On the contrary, the emphasis is on precedent, on the fact that such rights already existed and were sanctioned in the past, and that it was only a matter of reaffirming them solemnly and of providing adequate instruments for their defence. It took the American Revolution to convert these historic rights into natural rights. Undoubtedly the new spiritual climate and the rationalistic construction of the State which had found expression in Locke's writings had a share in this transformation. But an equal share was taken by the requirements of the political struggle. This no longer allowed the colonists, once they had detached themselves from the mother country, to invoke the 'rights of Englishmen', and thus led them almost necessarily to claim these rights on general grounds as 'rights of man'. The great Declaration of the French National Assembly of 1789 purported to define both the rights of man and those of the citizen; but no sooner had the storm of revolution settled, than the same rights reappeared as historic or positive rights, or rights of the citizen rather than of man. If this proves anything, it can prove only that the vindication of rights was a way of defining the boundaries between the State and the individual, and that negative liberty had become a constitutional principle, the basic constitutional principle of the modern State.

Nor is it less important, on the other hand, that there should always have been great differences of opinion in fixing these boundaries, and that these boundaries have never been so much called in question as at the present day. A case in point is the discussion still raging on the Continent about the existence and precise nature of a new category of rights and freedoms, the so-called 'social rights'.

It is not easy to define these new rights in a few words. In substance, they amount to the demand that all citizens should in fact be able to enjoy those freedoms which by right are equal for

all men. But a demand of this kind must in the end lead to the abandonment of the traditional idea of negative liberty as non-interference on the part of the State. Obviously, the promotion of social rights requires in the economic sphere not the abstention but the intervention of the State. This emerges clearly, for example, in the ruling of Article 3, § 2, of the Italian Constitution, which declares: 'It is the duty of the Republic to remove all obstacles of an economic and social nature, which by limiting in fact the liberty and equality of the citizens hinder the full development of human personality and the effective share of all workers in the political, economic, and social organisation of the country.'

The concept of social rights and freedom thus brings us back to the question, to what extent the liberal conception of the State is linked to a particular type of economic and social organisation. For there can be no doubt that it is difficult to reconcile this new concept of freedom with the interpretation of the liberal ideal which was given by the theorists of the bourgeois age, from Locke to Mill, and which inspired the constitutional Charters of the eighteenth and nineteenth centuries. This interpretation, as we saw, centred on the belief in the sanctity of private property and in the unsurpassed merits of full economic liberty. It is precisely that sanctity and those merits which are now called in question. They would have to be sacrificed if 'social justice' were to be realised. The conflict, to give it its proper name, is the conflict between Individualism and Socialism: and the problem of the survival of the liberal ideal is bound to be felt as particularly critical in societies which, like present-day Italian society, lie midway between the two. It is highly significant that two among the most distinguished leaders of Italian Liberalism should have been at loggerheads precisely on this point: the economist Einaudi maintained that Liberalism was inseparably tied to a *laissez-faire* structure of society, while the philosopher Croce maintained no less stoutly that liberal ideals and institutions could and should survive the radical social changes taking place in the modern world.

Faced with such a serious dilemma it is not enough to appeal to the experience of particular countries, like Britain for example, where the abandonment of economic Liberalism and the carrying through of courageous social reforms have apparently not weakened the old ideal of negative liberty, so well expressed in the saying 'An Englishman's home is his castle'. It is useful to go back to the crucial moment in the story of that liberty, to the time when it contributed decisively to the forming of the modern State in the shape of three basic demands: the removal of obstacles, the assurance of a sphere of individual independence and the confining of State action within well-known, precise limits. I believe that it would be possible to speak of a threat to that liberty which has inspired modern constitutions only if one of these three demands were rejected. As long as there is no return to the attribution of absolute and unlimited power to the State, as long as the life and independence of the citizen are not at the mercy of arbitrary decisions, in a word as long as no attack is made on those constitutional guarantees which form the bulwark of modern liberty, we cannot hold that certain actions of the State can be called illiberal. Least of all can such an accusation be brought against State action directed to removing obstacles to the full development of human personality, or to the effective participation of all citizens in political life, when such obstacles are due to their particular economic or social condition, or to the privileged position of certain groups within the State. On the contrary, such action would seem to me to be a continuance, a strengthening of that liberating process which, as I have tried to show, though often overlooked, is one of the highest merits of the modern State.

It was, after all, to this liberating task, to this removal of obstacles which constitutes negative liberty, that the Head of a great and powerful nation appealed during the last war. When F. D. Roosevelt (in his Message to Congress of 6 January 1941) proclaimed what came to be known as the Four Freedoms he was probably thinking only of offering a promise and hope to the peoples then crushed under the most horrible tyranny that history had ever witnessed. He was not outlining a political system, let alone a programme of social reform. But consider those freedoms in their constituent elements and in the historical perspective I have here outlined. The first and second – freedom of speech and expression, freedom to worship God each in his own way – are in the purest tradition of Liberalism; they might have been worded by John Stuart Mill. The fourth freedom, on the other hand – freedom from fear – expresses a much more elementary need; it almost evokes the spirit of Hobbes. As for the third freedom –

freedom from want – this too is a negative freedom, at least if my argument thus far has been correct. But surely nobody would venture to brand it as a bourgeois kind of freedom when it has inspired, and continues to inspire, those classes and peoples who have hitherto been disinherited, and who now claim their share of the common human heritage, just as, only a few centuries ago, did the ancestors of those classes and peoples who prefer today (fortunately they are only few in number) to shut themselves up in a sterile egoism, raving about 'natural inequalities' and non-existent privileges of 'race' or élite.

POSITIVE LIBERTY

Alexander d'Entrèves

THE concept of negative liberty, and the liberal theory of the State founded upon it, can be criticised for reasons different from, and better than, those which have been discussed in the preceding chapter. The point is whether this concept, in laying stress too exclusively on the limits of power, does not end by relegating to a secondary place, or at least by minimising, the problem of how power is to be exercised, and precisely by whom. The model of the liberal State was set up once and for all by Montesquieu, with his description of 'moderate governments' as the only ones where 'political liberty' can be found – the 'political liberty of the subject' being 'a tranquillity of mind, arising from the opinion each person has of his safety'. The choice of this model rules out any preference for a particular form of political organisation. It does not matter whether government is 'republican' or 'monarchical'. What matters is that power should not be 'abused', and this, as we know, can happen, according to Montesquieu, only where power is constitutionally limited. But if the test of a liberal government is that the rule of law should be secured and the area of private independence respected, strictly even an enlightened despot can satisfy that test. Frederick of Prussia was highly praised by Kant precisely on that score: for having enabled his subjects to use their reason and to speak their mind, thus removing all hindrances to the spread of enlightenment.

The claims advanced by modern man in the name of liberty are much more sweeping and complex. This is where we come up against what is usually described today as 'positive liberty', in relation to, if not in direct contrast with, negative liberty. The conceptual difference between these liberties is set forth in the following quotation from Sir Isaiah Berlin's recent Inaugural Lecture with a clarity I could hardly hope to improve upon:

> The answer to the question 'Who governs me?' is logically distinct from the question 'How far does government interfere with me?' It is in this difference that the great contrast between the two concepts of negative and positive liberty, in the end, consists. For the 'positive' sense of liberty comes to light if we try to answer the question, not 'What am I free to do or be?', but 'By whom am I ruled?' or 'Who is to say what I am, and what I am not, to be or do?' . . . The desire to be governed by myself, or at any rate to participate in the process by which my life is to be controlled, may be as deep a wish as that of a free area for action, and perhaps historically older. But it is not a desire for the same thing. . . . The 'positive' sense of the word 'liberty' derives from the wish on the part of the individual to be his own master.

The distinction so sharply drawn by Sir Isaiah is no new one. At the beginning of the nineteenth century Benjamin Constant laid it down equally forcibly in his celebrated essay *The Liberty of the Ancients compared with that of the Moderns*. It can even, without too much effort, be traced back to Aristotle. Its tremendous importance for political theory lies in the fact that it forces us to face the question which liberals often eschew, the question of the relation between liberty and form of government. Indeed, the very definition of positive liberty indicates how closely this notion is linked to that of democracy. If liberty consists in having a say in all decisions that affect us, then, as Aristotle pointed out, it can be realised only in a system of government based on equality, where all citizens share 'in the administration of justice and in the holding of office'. In a system such as this 'the masses must necessarily be sovereign'. Positive liberty and popular sovereignty go hand in hand. It might even be said – Cicero said it in different words long ago – that one is the ideological justification of the other. Democrats have always maintained that 'in no other city except in one where the people has the supreme power can liberty

find its abode: liberty, than which without doubt nothing is sweeter'.

We have already encountered the doctrine of popular sovereignty in another context. Much has been written on it and many theories have been put forward concerning its history, some of which have indeed gone too far in claiming the uninterrupted continuity of the doctrine ever since its first appearance in classical antiquity. Two points are of interest to our present inquiry: the assumption of popular sovereignty as the exclusive legitimation of power, and the assertion of a necessary and unbreakable bond between the concept of liberty and that of democracy. The clarification of these two points should help us to date with a certain precision the birth of the modern democratic State, or more exactly of the arguments which still lend it support.

As regards the first point, there can be little doubt that the claim put forward on behalf of popular sovereignty was closely associated, both logically and in time, with the emergence of those natural, original and indefeasible rights of the individual from which the liberal idea of the State also drew its origin. Locke's political theory provides the best proof of that close association. Once again, as had been the case with equality, a notion that was entirely new was linked with the past, and could be thought of as sprouting from the trunk of much older theories. Both the medieval idea of consent, and the Roman doctrine of the derivation of power from the people, certainly contributed to the shaping of the modern theory of popular sovereignty. But if the bottle was old, the wine was an entirely new one. What was new was the presentation of popular sovereignty as an 'original right' of the people, corresponding to the 'natural rights' of the individual. What was without precedent was the abstract construction of the State as the result of a social contract in which each individual contributes his share of rights and of power. Indeed, what took place was a revolution in political philosophy: it might be called a 'Copernican revolution', to paraphrase a famous image of Kant's. But it was a revolution in thought parallel to one in action, and it is futile to argue which of the two was prior in time and played the decisive role in shaping the first democratic régimes worthy to be so called in the Western hemisphere. What matters is that the leaders of those revolutions were men who not only believed in the democratic principle, in human equality, in the will of the people as the only legitimate source of power, but attempted to build a society based on those principles, and were ready to fight for them if need be. 'Really I think that the poorest hee that is in England hath a life to live as the greatest hee; and therefore truly, Sir, I thinke itt's cleare, that every man that is to live under a Government ought first by his owne consent to putt himself under that Government.' Colonel Rainborough's words are still remembered, as are those uttered by Lincoln on the battlefield of Gettysburg two centuries later, which also blazon forth a claim, an encouragement and a promise 'that government of the people, by the people, for the people, shall not perish from the earth'. Democracy so conceived is a commitment, a choice. To define it is to prescribe it. It is an ideal accepted or refused even before it can be justified on rational grounds.

Now among all the arguments which can be, and have been, advanced in defence of democracy the idea of positive liberty is by far the most attractive, the most convincing and at the same time the most ambiguous and the most liable to be misinterpreted or misunderstood. And Rousseau is of all political thinkers the one who has advanced that argument most cogently and drawn it with impeccable logic to its most extreme conclusions. For Rousseau the central problem of politics consisted in finding 'a form of association which will defend and protect with the whole common force the person and goods of each associate, and in which each, while uniting himself with all, may still obey himself alone, and remain as free as before'. 'This', Rousseau adds, 'is the fundamental problem of which the social contract provides the solution.' Through the social contract, with his entry into a 'civil state', man renounces his 'natural liberty', but only to find his true liberty, which is both 'civil' and 'moral', and consists in obedience to the rule of law: because man is free only when he is 'truly master of himself', and 'obedience to a law which we prescribe to ourselves is liberty'.

The notion of the social contract was thus stripped by Rousseau of any historic reference. It is, as Kant was to say, not an 'act' but an 'idea', the normative principle which allows us to conceive of political organisation as implying not the sacrifice of liberty but its attainment. Liberty in the positive sense is in fact nothing else than self-government, autonomy. It cannot be realised except when the power which commands is the selfsame power of him

who obeys. The achievement of liberty in the State is therefore conditional on the sovereignty of the 'general will'. 'Subjects' become 'citizens' only in as far as they share in that sovereignty and partake of the general will. Their complete surrender to the State, to the 'country' (*la patrie*), secures them 'against all personal dependence'. 'In this lies the key to the working of the political machine; this alone legitimises civil undertakings, which, without it, would be absurd, tyrannical and liable to the most frightful abuses.'

No one, probably, has ever gone as far as Rousseau in presenting democracy as the only legitimate government; certainly few before him had ever given the notion of positive liberty the prominence it is given in the *Social Contract*. This is the reason why it is always necessary to return to Rousseau when we want to understand the relevance of the idea of liberty to modern democratic thought. But it is also the reason why from the first, let alone in its later applications, Rousseau's argument has been under constant and bitter attack. Indeed, Rousseau, the greatest theorist of positive liberty, has seemed to many people to sacrifice Liberalism to Democracy. Early in the nineteenth century the liberal writer Constant denounced the *Social Contract* as 'the most terrible auxiliary of every form of despotism'. Not long ago Sir Isaiah Berlin, in a memorable broadcast, defined Rousseau as 'the most sinister and most formidable enemy of liberty in the whole history of modern thought'.

The liberty which was felt to be threatened by democracy of the Rousseauian type was first of all negative liberty, the liberty that was dearest to the heart of all liberals of the old school. The danger was conceived to spring from what was called 'egalitarian' democracy. This at any rate was the threat which the men of the nineteenth century had particularly in mind. They knew nothing as yet of the very different dangers which were to be revealed in times nearer our own. Above all, they saw the menace to liberty in what they called the 'tyranny of numbers', which they believed was the inevitable consequence of the principle of equality. Logically, this principle is certainly inseparable from the idea of democracy. Demanding, as it does, that everyone should share in the exercise of power, democracy was inevitably bound to lead to the progressive extension of the franchise, to the assumption that one vote counts as much as any other and to the need to solve eventual differences by having recourse to a simple numerical calculation, normally that of majority not further qualified. The fear of 'the despotism of the majority, which is nothing else than the right of the strongest', is a motif we find recurring in the writings of most early liberals, from Mallet du Pan to Constant and Mill. It was not dictated by any spirit of undemocratic reaction, but by a true concern for the proper working of liberal institutions and for the prevention of a new despotism, more exacting than the old.

A second threat which critics discovered in egalitarian democracy (the reference here is to constructive criticism only, not to that which was purely negative, and indulged in by reactionary writers) was that of the 'levelling of values', to call it by a modern name. This levelling was thought to be the inevitable result of the refusal to recognise the factual inequality of men, their different abilities, the role of self-differentiation in the dynamic of social life. The rejection of the aristocratic principle, some writers prophesied, would inevitably lead to a mass society with no standards left. Tocqueville is the nineteenth-century writer most clearly aware of this particular danger. He was, as he himself confessed, torn between his 'aristocratic instincts' and his 'rational belief' in democracy. His concern with the problem led him first to make use of his knowledge of democracy in America to produce a penetrating diagnosis of democratic experience in general. At a later stage it led him to trace the process of levelling in France to a time well before the Revolution, and to find its roots in the very structure of the *ancien régime* and in the rise of the modern bureaucratic, centralising State. Faced by the inevitable progress of egalitarianism, Tocqueville declared that he was smitten by a kind of 'religious terror'. Yet his mind remained cool, his conviction unshakable: 'The question is not how to reconstruct aristocratic society, but how to make liberty proceed out of that democratic state of society in which God has placed us.'

The last and worst danger which has been seen to arise from an unconditional application of the democratic principle is that it can provide the ground for nothing less than a final and total renunciation of liberty. Rousseau himself gave warning of this danger when he criticised certain theories of the social contract, and particularly that supported by Grotius. Grotius in fact, by presenting the contract as irrevocable, used that very notion to justify absolutism, or, as Rousseau puts it, 'to rob the peoples of

all their rights and invest kings with them by every conceivable artifice'. The history of the nineteenth century, to say nothing of our own, certainly does not lack cases where popular sovereignty was used as an instrument for the establishment of dictatorships. Plebiscites, referendums, direct appeals to the electorate, have served, as we know only too well, to provide both the legal and the ideological justification for régimes which proved to be far more absolute than all the absolute governments of past ages.

These are serious arguments which deserve to be pondered, all the more because they do, in most cases, accurately fit the facts. Experience still seems to confirm Tocqueville's remarks concerning the difference between the egalitarian, levelling, democracies of the Latin type, and those which have been called 'deferential democracies' of the English-speaking countries. In these countries, and especially in Britain – which unlike the Continent of Europe was spared the great social upheaval of the French Revolution – a foreign observer is struck by what can only be described as the survival of many aristocratic elements in the structure of society. Such elements, together with the frame of mind which they foster, act as a corrective to the rigid logic of democratic egalitarianism, though it may well be that the most valid corrective lies in the spirit of independence that has long prevailed on these shores. However that may be, the case of Britain does seem to prove the possibility of reconciling positive and negative liberty, democracy and liberalism. Not only does the safeguard of individual rights, and of an area of free individual development, set a limit to the 'tyranny of numbers'. That very safeguard is a condition of the proper working of democracy: without negative liberty, the attainment of positive liberty, of real and not illusory self-government, will be retarded, if not impeded, throughout.

For indeed, if democracy is to work properly, obstacles must be removed that would prevent men from thinking and judging freely. They must be given the chance of discussing and weighing up their leaders before confirming by their consent the power of the so-called élites. But for this to happen, the free play of political forces must be assured in the State. The principle of equality must be truly respected by guaranteeing to minorities the possibility of becoming majorities. Above all, we must never give to the decisions of majorities more than a purely pragmatic value, remembering, as the saying goes, that the only reason for counting heads is that it is easier to count them than to cut them off. Those decisions are human and fallible. They cannot entail a duty to submit to them as to an indisputable and ultimate truth. Subject to all these cautions, as is the case in modern constitutions, egalitarian democracy must not be confused with another, very different kind of democracy, which has only lately risen on the political horizon: 'totalitarian' democracy – the most terrible threat, in the opinion of many, to the ideal of liberty on which rests the modern notion of the State.

The key to this kind of democracy can undoubtedly be found once again in Rousseau, and in the ambivalence of that concept of positive liberty which, as we saw, plays a leading role in his construction of the State. The reason for this ambivalence and for the diverging interpretations which have been given of Rousseau's political thought can in turn be traced back to the ambiguity of his language, to what I should be inclined to call his skilful manipulation of words. Whether that manipulation was intentional is open to question: the fact remains that there is something deliberately obscure in Rousseau's presentation of his two central concepts, the social contract and the general will. The social contract is spoken of almost in terms of a mystical experience. It demands the 'total alienation' of the individual, offering in exchange a kind of rebirth, which will turn him from 'a stupid and unimaginative animal' into 'an intelligent being and a man'. Even more baffling is Rousseau's conception of the general will, which has called forth an endless literature. Clearly, the general will is for Rousseau not merely the rightful holder of sovereignty. It embodies a supreme ethical value: for it alone 'is always right', and only in obedience to it does man find the fulfilment of his moral life. For this very reason the general will does not and cannot for Rousseau coincide with 'the will of all', at any rate not without further qualification. The will of all can in fact express particular and variable interests, whereas the general will is 'always constant, unalterable and pure'. It does not show itself 'in long debates, discussions and tumult'; rather, it is best revealed in unanimity, in the greatest possible concert of minds. It is therefore necessary to put an end to the 'harmful divisions' due to 'parties' and 'sects' (Rousseau expressly borrows these words from Machiavelli). 'There should be no partial society within the State', in order that 'the people shall in no way deceive itself' and nothing may disturb

the mysterious revelation of truth in each citizen's mind, thinking 'only his own thoughts'.

Nor is this all. The 'blind multitude', says Rousseau, needs to be 'guided'.

> Of itself the people wills always the good, but of itself it by no means always sees it. The general will is always in the right, but the judgement which guides it is not always enlightened. It must be got to see objects as they are, and sometimes as they ought to appear to it; it must be shown the good road it is in search of. . . . The individuals see the good they reject; the public wills the good it does not see. All stand equally in need of guidance.

Rousseau leaves us in no doubt as to what he means by such guidance. He states in *Émile* that the best institutions are those which best succeed in 'denaturing' man. He repeats in the *Social Contract* that the shaping of a nation may require that human nature itself be changed. Such a formidable undertaking can be achieved only by training and by compulsion: souls must be taught as well as forced to fly. 'Whoever refuses to obey the general will shall be compelled to do so by the whole body. This means nothing less than that he will be forced to be free.' Indeed, the general will may appear to run contrary to my convictions and to my vote. But 'this proves neither more nor less than that I was mistaken and that what I thought to be the general will was not so. If my particular opinion had carried the day, I should have achieved the opposite of what was my will; and it is in that case that I should not have been free.'

The need of guidance, the wrongness of dissent; indoctrination, self-criticism, re-education: all the paraphernalia of this century's totalitarianism seem, alas, to be already contained in these statements of Rousseau. One can well understand why, in search of culprits, many respectable modern scholars should have seized gladly upon Rousseau, using him as a scapegoat for all our troubles as he had already been used more than a century and a half before. I believe that such views contain a certain amount of exaggeration and distortion; but I cannot here enter into a discussion both of what might be called the case for Rousseau's prosecution and of that for his defence. One point, at any rate, must be made: viz. that with Rousseau begins that manipulation of words I have already mentioned, which has led to such widely contrasting concepts of liberty and democracy as dominate our present world. What happened would seem to be that words belonging to the traditional vocabulary of politics were emptied of their content, and new meanings were substituted, which proved eventually to be quite the opposite of the old. It would, of course, be easy to produce countless examples of this kind of manipulation in our own time, simply by turning to current political propaganda. But it is only fair to recognise that the earliest and largest part of the responsibility for what has happened lies with political philosophers, and precisely with those who, starting with Rousseau and proceeding with Hegel, shared in the greatest manipulation of all: that which changed the liberal formula of the State as an 'instrument' of liberty into the totalitarian slogan of the State as the 'embodiment' of freedom. It is impossible not to feel perplexed, for example, on reading in De Ruggiero's *History of European Liberalism* – a book which is still very popular, even outside Italy, and whose author was without doubt a sincere believer in Liberalism – a statement like the following: 'We are today so much accustomed to the idea of the liberal State that we do not notice its paradoxical character, which was plain enough to its first inexperienced observers. The State, the organ of coercion *par excellence*, has become the highest expression of liberty.'

For my part, besides noticing the confusion which Hegelianism can foster even in a clear-sighted Latin mind, I believe that the only possible answer to this sort of argument is that, contrary to De Ruggiero's opinion, the first observers were not so far from the truth, or at least that they were reasoning correctly in terms of ordinary language and according to common sense. As with the opposition between force and consent, no dialectical artifice has ever succeeded in overcoming the opposition between coercion and liberty. 'To force men to be free' may seem to some people to be the greatest oracular utterance ever made by political philosophy. To my generation this sentence can only be a painful reminder of other similar phrases which could be read at the entrance to concentration camps. Better to give things their true names, to recognise facts for what they are. 'The organ of coercion *par excellence*' was certainly never in the past 'the expression of liberty', and need not be so today. There can and do exist democracies which deny individual liberty, just as there can be 'liberal' States where only a small number of citizens have any effective share of

power. The State is, and can become, the expression of liberty only on condition that certain 'rules of the game' are respected. Negative and positive liberty constitute these rules, and this and nothing else is meant by saying that the purpose of power is to secure human freedom.

But if this is how things stand, then it is clear that we have not yet come to the end of our labours. Besides demanding respect for the rules, there may well be a question of asking what is at stake in the game: metaphor apart, it is a question of seeing whether it is possible to determine not only the form but the content of State action, to establish, as was long believed possible, a firm and objective standard – the standard of the 'common good' – for the final legitimation of power. Writers with every kind of background have shared this belief. No better conclusion to our long journey could be found than a pause to reflect upon this venerable notion.

An isolated individual can consume only what he produces for himself; like Robinson Crusoe, he must turn his hand to many different activities, however unskilled he may be in some of them. A whole community, on the other hand, can practise a division of labour, under which each member specialises in a single line of production: between them, they will produce far more than an equivalent number of Crusoes. Thus, one of the advantages of living in society is that its members can enjoy a higher standard of living than if they remained outside it. The first extract from Adam Smith's 'The Wealth of Nations' (first published in 1776) is a classic exposition of the principle of specialisation.

In a society whose members are free to buy and sell as they choose, the desire for private gain will direct resources into the occupations in which they are most productive; without any plan or control on the part of government, the forces of supply and demand will cause wants to be satisfied to the largest extent possible given the total resources available. Such, at any rate, was the belief of the economists who followed Adam Smith, a statement of whose position forms the second extract.

Specialisation means that each individual produces far more than he needs of the single commodity in which he specialises, and exchanges the surplus for the output of other specialised producers; without the possibility of exchange, specialisation would not be practicable. Adam Smith asserted the existence of 'a certain propensity in human nature . . . to truck, barter, and exchange one thing for another', which 'originally gives occasion to the division of labour'. It can be seen at work in the third extract, which describes the economic relationships which arise when a number of people are thrown together in the primitive conditions of a prisoner-of-war camp.

THE DIVISION OF LABOUR

Adam Smith

THE greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity and judgement with which it is anywhere directed, or applied, seem to have been the effects of the division of labour. . . .

To take an example, therefore, from a very trifling manufacture; but one in which the division of labour has been very often taken notice of, the trade of the pin-maker; a workman not educated to this business, nor acquainted with the use of the machinery employed in it, could scarce, perhaps, with his utmost industry, make one pin in a day, and certainly could not make twenty. But in the way in which this business is now carried on, not only the whole work is a peculiar trade, but it is divided into a number of branches, of which the greater part are likewise peculiar trades. One man draws out the wire, another straights it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them. I have seen a small manufactory of this kind where ten men only were employed, and where some of them consequently performed two or three distinct operations. But though they were very poor, and therefore but indifferently accommodated with the necessary machinery, they could, when they exerted themselves, make among them about twelve pounds of pins a day. There are in a pound upwards of four thousand pins of a middling size. Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made twenty, perhaps not one pin in a day; that is, certainly, not the two hundred and fortieth, perhaps not the four thousand eight hundredth part of

what they are at present capable of performing, in consequence of a proper division and combination of their different operations. . . .

This great increase of the quantity of work which, in consequence of the division of labour, the same number of people are capable of performing, is owing to three different circumstances; first, to the increase of dexterity in every particular workman; secondly, to the saving of the time which is commonly lost in passing from one species of work to another; and lastly, to the invention of a great number of machines which facilitate and abridge labour, and enable one man to do the work of many. . . .

THE 'INVISIBLE HAND'

Adam Smith

THE produce of industry is what it adds to the subject or materials on which it is employed. In proportion as the value of this produce is great or small, so will likewise be the profits of the employer. But it is only for the sake of profit that any man employs a capital in the support of industry; and he will always, therefore, endeavour to employ it in the support of that industry of which the produce is likely to be of the greatest value, or to exchange for the greatest quantity either of money or of other goods.

But the annual revenue of every society is always precisely equal to the exchangeable value of the whole annual produce of its industry, or rather is precisely the same thing with that exchangeable value. As every individual, therefore, endeavours as much as he can both to employ his capital in the support of domestic industry, and so to direct that industry that its produce may be of the greatest value; every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. . . . By directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. I have never known much good done by those who affected to trade for the public interest.

THE ECONOMIC ORGANISATION OF A P.O.W. CAMP

R. A. Radford

(The author was a British prisoner in Germany during World War II.)

AFTER allowance has been made for abnormal circumstances, the social institutions, ideas and habits of groups in the outside world are to be found reflected in a prisoner-of-war camp. One aspect of social organisation is to be found in economic activity, and this, along with other manifestations of a group existence, is to be found in any P.O.W. camp. True, a prisoner is not dependent on his exertions for the provision of the necessaries, or even the luxuries of life, but through his economic activity, the exchange of goods and services, his standard of material comfort is considerably enhanced. And this is a serious matter to the prisoner: he is not 'playing at shops', even though the small scale of the transactions and the simple expression of comfort and wants in terms of cigarettes and jam, razor blades and writing paper, make the urgency of those needs difficult to appreciate, even by an ex-prisoner of some three months' standing.

Nevertheless, it cannot be too strongly emphasised that economic activities do not bulk so large in prison society as they do in the larger world. There can be little production; the emphasis lies in exchange and the media of exchange.

But it would be wrong to underestimate the importance of economic activity. Everyone receives a roughly equal share of essentials; it is by trade that individual preferences are given expression and comfort increased. All at some time, and most people regularly, make exchanges of one sort or another.

Between individuals there was active trading in all consumer goods and in some services. Most trading was for food against cigarettes or other foodstuffs, but cigarettes rose from the status

of a normal commodity to that of currency. Reichsmarks existed but had no circulation save for gambling debts, as few articles could be purchased with them from the canteen.

Our supplies consisted of rations provided by the detaining power and (principally) the contents of Red Cross food parcels – tinned milk, jam, butter, biscuits, bully, chocolate, sugar, etc., and cigarettes. So far the supplies to each person were equal and regular. Private parcels of clothing, toilet requisites and cigarettes were also received, and here equality ceased owing to the different numbers dispatched and the vagaries of the post. All these articles were the subject of trade and exchange.

THE DEVELOPMENT AND ORGANISATION OF THE MARKET

Very soon after capture people realised that it was both undesirable and unnecessary, in view of the limited size and the equality of supplies, to give away or to accept gifts of cigarettes or food. 'Goodwill' developed into trading as a more equitable means of maximising individual satisfaction.

We reached a transit camp in Italy about a fortnight after capture and received one-quarter of a Red Cross food parcel each a week later. At once exchanges, already established, multiplied in volume. Starting with simple direct barter, such as a non-smoker giving a smoker friend his cigarette issue in exchange for a chocolate ration, more complex exchanges soon became an accepted custom. Stories circulated of a padre who started off round the camp with a tin of cheese and five cigarettes and returned to his bed with a complete parcel in addition to his original cheese and cigarettes; the market was not yet perfect. Within a week or two, as the volume of trade grew, rough scales of exchange values came into existence. Sikhs, who had at first exchanged tinned beef for practically any other foodstuff, began to insist on jam and margarine. It was realised that a tin of jam was worth ½ lb of margarine plus something else, that a cigarette issue was worth several chocolate issues, and a tin of diced carrots was worth practically nothing.

In this camp we did not visit other bungalows very much and prices varied from place to place; hence the germ of truth in the story of the itinerant priest. By the end of a month, when we reached our permanent camp, there was a lively trade in all commodities and their relative values were well known, and expressed not in terms of one another – one didn't quote bully in terms of sugar – but in terms of cigarettes. The cigarette became the standard of value. In the permanent camp people started by wandering through the bungalows calling their offers – 'cheese for seven' (cigarettes) – and the hours after parcel issue were Bedlam. The inconveniences of this system soon led to its replacement by an Exchange and Mart notice board in every bungalow, where under the headings 'name', 'room number', 'wanted' and 'offered' sales and wants were advertised. When a deal went through, it was crossed off the board. The public and semi-permanent records of transactions led to cigarette prices being well known and thus tending to equality throughout the camp, although there were always opportunities for an astute trader to make a profit from arbitrage. With this development everyone, including non-smokers, was willing to sell for cigarettes, using them to buy at another time and place. Cigarettes became the normal currency, though, of course, barter was never extinguished.

The unity of the market and the prevalence of a single price varied directly with the general level of organisation and comfort in the camp. A transit camp was always chaotic and uncomfortable: people were overcrowded, no one knew where anyone else was living, and few took the trouble to find out. Organisation was too slender to include an Exchange and Mart board, and private advertisements were the most that appeared. Consequently a transit camp was not one market but many. The price of a tin of salmon is known to have varied by two cigarettes in twenty between one end of a hut and the other. Despite a high level of organisation in Italy, the market was morcellated in this manner at the first transit camp we reached after our removal to Germany in the autumn of 1943. In this camp – Stalag VII A at Moosburg in Bavaria – there were up to 50,000 prisoners of all nationalities. French, Russians, Italians and Jugoslavs were free to move about within the camp; British and Americans were confined to their compounds, although a few cigarettes given to a sentry would always procure permission for one or two men to visit other compounds. The people who first visited the highly organised French trading centre, with its stalls and known prices, found coffee extract – relatively cheap among the tea-drinking English –

commanding a fancy price in biscuits or cigarettes, and some enterprising people made small fortunes that way. (Incidentally we found out later that much of the coffee went 'over the wire' and sold for phenomenal prices at black-market cafés in Munich: some of the French prisoners were said to have made substantial sums in Reichsmarks. This was one of the few occasions on which our normally closed economy came into contact with other economic worlds.)

Eventually public opinion grew hostile to these monopoly profits – not everyone could make contact with the French – and trading with them was put on a regulated basis. Each group of beds was given a quota of articles to offer and the transaction was carried out by accredited representatives from the British compound, with monopoly rights. The same method was used for trading with sentries elsewhere, as in this trade secrecy and reasonable prices had a peculiar importance, but as is ever the case with regulated companies, the interloper proved too strong.

The permanent camps in Germany saw the highest level of commercial organisation. In addition to the Exchange and Mart notice boards, a shop was organised as a public utility, controlled by representatives of the Senior British Officer, on a no-profit basis. People left their surplus clothing, toilet requisites and food there until they were sold at a fixed price in cigarettes. Only sales in cigarettes were accepted – there was no barter – and there was no higgling. For food at least there were standard prices; clothing is less homogeneous and the price was decided around a norm by the seller and the shop manager in agreement; shirts would average say 80, ranging from 60 to 120 according to quality and age. Of food, the shop carried small stocks for convenience; the capital was provided by a loan from the bulk store of Red Cross cigarettes and repaid by a small commission taken on the first transactions. Thus the cigarette attained its fullest currency status, and the market was almost completely unified.

It is thus to be seen that a market came into existence without labour or production. Despite a roughly equal distribution of resources, a market came into spontaneous operation, and prices were fixed by the operation of supply and demand. It is difficult to reconcile this fact with the labour theory of value.

Actually there was an embryo labour market. Even when cigarettes were not scarce, there was usually some unlucky person willing to perform services for them. Laundrymen advertised at two cigarettes a garment. Battledress was scrubbed and pressed and a pair of trousers lent for the interim period for twelve. A good pastel portrait cost thirty or a tin of 'Kam'. Odd tailoring and other jobs similarly had their places.

There were also entrepreneurial services. There was a coffee-stall owner who sold tea, coffee or cocoa at two cigarettes a cup, buying his raw materials at market prices and hiring labour to gather fuel and to stoke; he actually enjoyed the services of a chartered accountant at one stage. After a period of great prosperity he overreached himself and failed disastrously for several hundred cigarettes. Such large-scale private enterprise was rare, but several middlemen or professional traders existed. The padre in Italy, or the men at Moosburg who opened trading relations with the French, are examples: the more subdivided the market, the less perfect the advertisement of prices, and the less stable the prices, the greater was the scope for these operators. One man capitalised his knowledge of Urdu by buying meat from the Sikhs and selling butter and jam in return; as his operations became better known more and more people entered this trade, and prices in the Indian Wing approximated more nearly to those elsewhere, though to the end a 'contact' among the Indians was valuable, as linguistic difficulties prevented the trade from being quite free. Some were specialists in the Indian trade, the food, clothing or even the watch trade. Middlemen traded on their own account or on commission. Price-fixing agreements were suspected and the traders certainly co-operated. Nor did they welcome newcomers. Unfortunately the writer knows little of the workings of these people: public opinion was hostile and the professionals were usually of a retiring disposition.

One trader in food and cigarettes, operating in a period of dearth, enjoyed a high reputation. His capital, carefully saved, was originally about fifty cigarettes, with which he bought rations on issue days and held them until the price rose just before the next issue. He also picked up a little by arbitrage; several times a day he visited every Exchange and Mart notice board and took advantage of every discrepancy between prices and goods offered and wanted. His knowledge of prices, billets and names of those who had received cigarette parcels was phenomenal. By these

means he kept himself smoking steadily – his profits – while his capital remained intact.

Sugar was issued on Saturday. About Tuesday two of us used to visit Sam and make a deal; as old customers he would advance as much of the price as he could spare then, and entered the transaction in a book. On Saturday morning he left cocoa tins on our beds for the ration, and picked them up on Saturday afternoon. We were hoping for a calendar at Christmas, but Sam failed too. He was left holding a big black treacle issue when the price fell, and in this weakened state was unable to withstand an unexpected arrival of parcels and the consequent price fluctuations. He paid in full, but from his capital. The next Tuesday, when I paid my usual visit, he was out of business.

Credit entered into many, perhaps into most transactions, in one form or another. Sam paid in advance as a rule for his purchases of future deliveries of sugar, but many buyers asked for credit, whether the commodity was sold spot or future. Naturally prices varied according to the terms of sale. A treacle ration might be advertised for four cigarettes now or five next week. And in the future market 'bread now' was a vastly different thing from 'bread Thursday'. Bread was issued on Thursday and Monday, four and three days' rations respectively, and by Wednesday and Sunday night it had risen at least one cigarette per ration, from seven to eight, by supper time. One man always saved a ration to sell then at the peak price: his offer of 'bread now' stood out on the board among a number of 'bread Monday's' fetching one or two less, or not selling at all – and he always smoked on Sunday night.

THE CIGARETTE CURRENCY

Although cigarettes as currency exhibited certain peculiarities, they performed all the functions of a metallic currency as a unit of account, as a measure of value and a store of value, and shared most of its characteristics. They were homogeneous, reasonably durable and of convenient size for the smallest or, in packets, for the largest transactions. Incidentally, they could be clipped or sweated by rolling them between the fingers so that tobacco fell out.

Cigarettes were also subject to the working of Gresham's Law. Certain brands were more popular than others as smokes, but for currency purposes a cigarette was a cigarette. Consequently buyers used the poorer qualities and the Shop rarely saw the more popular brands: cigarettes such as Churchman's No. 1 were rarely used for trading. At one time cigarettes hand-rolled from pipe tobacco began to circulate. Pipe tobacco was issued in lieu of cigarettes by the Red Cross at a rate of twenty-five cigarettes to the ounce, and this rate was standard in exchanges, but an ounce would produce thirty home-made cigarettes. Naturally, people with machine-made cigarettes broke them down and re-rolled the tobacco, and the real cigarette virtually disappeared from the market. Hand-rolled cigarettes were not homogeneous and prices could no longer be quoted in them with safety: each cigarette was examined before it was accepted and thin ones were rejected, or extra demanded as a make-weight. For a time we suffered all the inconveniences of a debased currency.

Machine-made cigarettes were always universally acceptable, both for what they would buy and for themselves. It was this intrinsic value which gave rise to their principal disadvantage as currency, a disadvantage which exists, but to a far smaller extent, in the case of metallic currency; that is, a strong demand for non-monetary purposes. Consequently our economy was repeatedly subject to deflation and to periods of monetary stringency. While the Red Cross issue of fifty or twenty-five cigarettes per man per week came in regularly, and while there were fair stocks held, the cigarette currency suited its purpose admirably. But when the issue was interrupted, stocks soon ran out, prices fell, trading declined in volume and became increasingly a matter of barter. This deflationary tendency was periodically offset by the sudden injection of new currency. Private cigarette parcels arrived in a trickle throughout the year, but the big numbers came in quarterly when the Red Cross received its allocation of transport. Several hundred thousand cigarettes might arrive in the space of a fortnight. Prices soared, and then began to fall, slowly at first but with increasing rapidity as stocks ran out, until the next big delivery. Most of our economic troubles could be attributed to this fundamental instability.

PRICE MOVEMENTS

Many factors affected prices, the strongest and most noticeable being the periodical currency inflation and deflation described in the last paragraphs. The periodicity of this price cycle depended on cigarette and, to a far lesser extent, on food deliveries. At one time in the early days, before any private parcels had arrived and when there were no individual stocks, the weekly issue of cigarettes and food parcels occurred on a Monday. The non-monetary demand for cigarettes was great, and less elastic than the demand for food; consequently prices fluctuated weekly, falling towards Sunday night and rising sharply on Monday morning. Later, when many people held reserves, the weekly issue had no such effect, being too small a proportion of the total available. Credit allowed people with no reserves to meet their non-monetary demand over the weekend.

The general price level was affected by other factors. An influx of new prisoners, proverbially hungry, raised it. Heavy air raids in the vicinity of the camp probably increased the non-monetary demand for cigarettes and accentuated deflation. Good and bad war news certainly had its effect, and the general waves of optimism and pessimism which swept the camp were reflected in prices. Before breakfast one morning in March 1945, a rumour of the arrival of parcels and cigarettes was circulated. Within ten minutes I sold a treacle ration, for four cigarettes (hitherto offered in vain for three), and many similar deals went through. By ten o'clock the rumour was denied, and treacle that day found no more buyers even at two cigarettes.

More interesting than changes in the general price level were changes in the price structure. Changes in the supply of a commodity, in the German ration scale or in the make-up of Red Cross parcels, would raise the price of one commodity relative to others. Tins of oatmeal, once a rare and much sought-after luxury in the parcels, became a commonplace in 1943, and the price fell. In hot weather the demand for cocoa fell and that for soap rose. A new recipe would be reflected in the price level: the discovery that raisins and sugar could be turned into an alcoholic liquor of remarkable potency reacted permanently on the dried-fruit market. The invention of electric immersion heaters run off the power points made tea, a drug on the market in Italy, a certain seller in Germany.

In August 1944 the supplies of parcels and cigarettes were both halved. Since both sides of the equation were changed in the same degree, changes in prices were not anticipated. But this was not the case: the non-monetary demand for cigarettes was less elastic than the demand for food, and food prices fell a little. More important, however, were the changes in the price structure. German margarine and jam, hitherto valueless owing to adequate supplies of Canadian butter and marmalade, acquired a new value. Chocolate, popular and a certain seller, and sugar fell. Bread rose; several standing contracts of bread for cigarettes were broken, especially when the bread ration was reduced a few weeks later.

In February 1945 the German soldier who drove the ration wagon was found to be willing to exchange loaves of bread at the rate of one loaf for a bar of chocolate. Those in the know began selling bread and buying chocolate, by then almost unsaleable in a period of serious deflation. Bread, at about 40, fell slightly; chocolate rose from 15; the supply of bread was not enough for the two commodities to reach parity, but the tendency was unmistakable.

The substitution of German margarine for Canadian butter when parcels were halved naturally affected their relative values, margarine appreciating at the expense of butter. Similarly two brands of dried milk, hitherto differing in quality and therefore in price by five cigarettes a tin, came together in price as the wider substitution of the cheaper raised its relative value.

Enough has been cited to show that any change in conditions affected both the general price level and the price structure. It was this latter phenomenon which wrecked our planned economy.

PAPER CURRENCY: BULLY MARKS

Around D-Day, food and cigarettes were plentiful, business was brisk and the camp in an optimistic mood. Consequently the Entertainments Committee felt the moment opportune to launch a restaurant, where food and hot drinks were sold while a band and variety turns performed. Earlier experiments, both public

and private, had pointed the way, and the scheme was a great success. Food was bought at market prices to provide the meals and the small profits were devoted to a reserve fund and used to bribe Germans to provide grease-paints and other necessities for the camp theatre. Originally meals were sold for cigarettes, but this meant that the whole scheme was vulnerable to the periodic deflationary waves, and furthermore heavy smokers were unlikely to attend much. The whole success of the scheme depended on an adequate amount of food being offered for sale in the normal manner.

To increase and facilitate trade, and to stimulate supplies and customers therefore, and secondarily to avoid the worst effects of deflation when it should come, a paper currency was organised by the Restaurant and the Shop. The Shop bought food on behalf of the Restaurant with paper notes and the paper was accepted equally with the cigarettes in the Restaurant or Shop, and passed back to the Shop to purchase more food. The Shop acted as a bank of issue. The paper money was backed 100 per cent by food; hence its name, the Bully Mark (BMk). There could be no over-issues, as is permissible with a normal bank of issue, since the eventual dispersal of the camp and consequent redemption of all BMks was anticipated in the near future.

Originally one BMk was worth one cigarette, and for a short time both circulated freely inside and outside the Restaurant. Prices were quoted in BMks and cigarettes with equal freedom – and for a short time the BMk showed signs of replacing the cigarette as currency. The BMk was tied to food, but not to cigarettes; as it was issued against food, say 45 for a tin of milk and so on, any reduction in the BMk prices of food would have meant that there were unbacked BMks in circulation. But the price of both food and BMks could and did fluctuate with the supply of cigarettes.

While the Restaurant flourished, the scheme was a success; the Restaurant bought heavily, all foods were saleable and prices were stable.

In August parcels and cigarettes were halved and the camp was bombed. The Restaurant closed for a short while and sales of food became difficult. Even when the Restaurant reopened, the food and cigarette shortage became increasingly acute and people were unwilling to convert such valuable goods into paper and to hold them for luxuries like snacks and tea. Less of the right kinds of food for the Restaurant were sold, and the Shop became glutted with dried fruit, chocolate, sugar, etc., which the Restaurant could not buy. The price level and the price structure changed. The BMk fell to four-fifths of a cigarette and eventually farther still, and it became unacceptable save in the Restaurant. There was a flight from the BMk, no longer convertible into cigarettes or popular foods. The cigarette re-established itself.

But the BMk was sound! The Restaurant closed in the New Year with a progressive food shortage and the long evenings without lights due to intensified Allied air raids, and BMks could only be spent in the Coffee Bar – relict of the Restaurant – or on the few unpopular foods in the Shop, the owners of which were prepared to accept them. In the end all holders of BMks were paid in full, in cups of coffee or in prunes. People who had bought BMks for cigarettes or valuable jam or biscuits in their heyday were aggrieved that they should have stood the loss involved by their restricted choice, but they suffered no actual loss of market value.

PRICE FIXING

Along with this scheme came a determined attempt at a planned economy, at price fixing. The Medical Officer had long been anxious to control food sales, for fear of some people selling too much, to the detriment of their health. The deflationary waves and their effects on prices were inconvenient to all and would be dangerous to the Restaurant which had to carry stocks. Furthermore, unless the BMk was convertible into cigarettes at about par, it had little chance of gaining confidence and of succeeding as a currency. As has been explained, the BMk was tied to food but could not be tied to cigarettes, which fluctuated in value. Hence, while BMk prices of food were fixed for all time, cigarette prices of food and BMks varied.

The Shop, backed by the Senior British Officer, was now in a position to enforce price control both inside and outside its walls. Hitherto a standard price had been fixed for food left for sale in the Shop, and prices outside were roughly in conformity with this scale, which was recommended as a 'guide' to sellers, but fluc-

tuated a good deal around it. Sales in the Shop at recommended prices were apt to be slow, though a good price might be obtained; sales outside could be made more quickly at lower prices. (If sales outside were to be at higher prices, goods were withdrawn from the Shop until the recommended price rose; but the recommended price was sluggish and could not follow the market closely by reason of its very purpose, which was stability.) The Exchange and Mart notice boards came under the control of the Shop: advertisements which exceeded a 5 per cent departure from the recommended scale were liable to be crossed out by authority; unauthorised sales were discouraged by authority and also by public opinion, strongly in favour of a just and stable price. (Recommended prices were fixed partly from market data, partly on the advice of the Medical Officer.)

At first the recommended scale was a success: the Restaurant, a big buyer, kept prices stable around this level; opinion and the 5 per cent tolerance helped. But when the price level fell with the August cuts and the price structure changed, the recommended scale was too rigid. Unchanged at first, as no deflation was expected, the scale was tardily lowered, but the prices of goods on the new scale remained in the same relation to one another, owing to the BMk, while on the market the price structure had changed. And the modifying influence of the Restaurant had gone. The scale was moved up and down several times, slowly following the inflationary and deflationary waves, but it was rarely adjusted to changes in the price structure. More and more advertisements were crossed off the board, and black-market sales at unauthorised prices increased; eventually public opinion turned against the recommended scale and authority gave up the struggle. In the last few weeks, with unparalleled deflation, prices fell with alarming rapidity, no scales existed, and supply and demand, alone and unmellowed, determined prices.

PUBLIC OPINION

Public opinion on the subject of trading was vocal if confused and changeable, and generalisations as to its direction are difficult and dangerous. A tiny minority held that all trading was undesirable as it engendered an unsavoury atmosphere; occasional frauds and sharp practices were cited as proof. Certain forms of trading were more generally condemned; trade with the Germans was criticised by many. Red Cross toilet articles, which were in short supply and only issued in cases of actual need, were excluded from trade by law and opinion working in unshakeable harmony. At one time, when there had been several cases of malnutrition reported among the more devoted smokers, no trade in German rations was permitted, as the victims became an additional burden on the depleted food reserves of the Hospital. But while certain activities were condemned as anti-social, trade itself was practised, and its utility appreciated, by almost everyone in the camp.

More interesting was opinion on middlemen and prices. Taken as a whole, opinion was hostile to the middleman. His function, and his hard work in bringing buyer and seller together, were ignored; profits were not regarded as a reward for labour, but as the result of sharp practices. Despite the fact that his very existence was proof to the contrary, the middleman was held to be redundant in view of the existence of an official Shop and the Exchange and Mart. Appreciation only came his way when he was willing to advance the price of a sugar ration, or to buy goods spot and carry them against a future sale. In these cases the element of risk was obvious to all, and the convenience of the service was felt to merit some reward. Particularly unpopular was the middleman with an element of monopoly, the man who contacted the ration-wagon driver, or the man who utilised his knowledge of Urdu. And middlemen as a group were blamed for reducing prices. Opinion notwithstanding, most people dealt with a middleman, whether consciously or unconsciously, at some time or another.

There was a strong feeling that everything had its 'just price' in cigarettes. While the assessment of the just price, which incidentally varied between camps, was impossible of explanation, this price was nevertheless pretty closely known. It can best be defined as the price usually fetched by an article in good times when cigarettes were plentiful. The 'just price' changed slowly; it was unaffected by short-term variations in supply, and while opinion might be resigned to departures from the 'just price', a strong feeling of resentment persisted. A more satisfactory definition of the 'just price' is impossible. Everyone knew what it was, though no one could explain why it should be so.

As soon as prices began to fall with a cigarette shortage, a clam-

our arose, particularly against those who held reserves and who bought at reduced prices. Sellers at cut prices were criticised and their activities referred to as the black market. In every period of dearth the explosive question of 'Should non-smokers receive a cigarette ration?' was discussed to profitless length. Unfortunately it was the non-smoker, or the light smoker with his reserves, along with the hated middleman, who weathered the storm most easily.

The popularity of the price-fixing scheme, and such success as it enjoyed, were undoubtedly the result of this body of opinion. On several occasions the fall of prices was delayed by the general support given to the recommended scale. The onset of deflation was marked by a period of sluggish trade; prices stayed up but no one bought. Then prices fell on the black market, and the volume of trade revived in that quarter. Even when the recommended scale was revised, the volume of trade in the Shop would remain low. Opinion was always overruled by the hard facts of the market.

Curious arguments were advanced to justify price fixing. The recommended prices were in some way related to the calorific values of the foods offered: hence some were overvalued and never sold at these prices. One argument ran as follows: not everyone has private cigarette parcels; thus, when prices were high and trade good in the summer of 1944, only the lucky rich could buy. This was unfair to the man with few cigarettes. When prices fell in the following winter, prices should be pegged high so that the rich, who had enjoyed life in the summer, should put many cigarettes into circulation. The fact that those who sold to the rich in the summer had also enjoyed life then, and the fact that in the winter there was always someone willing to sell at low prices, were ignored. Such arguments were hotly debated each night after the approach of Allied aircraft extinguished all lights at 8 p.m. But prices moved with the supply of cigarettes, and refused to stay fixed in accordance with a theory of ethics.

CONCLUSION

The economic organisation described was both elaborate and smooth-working in the summer of 1944. Then came the August cuts and deflation. Prices fell, rallied with deliveries of cigarette parcels in September and December, and fell again. In January 1945 supplies of Red Cross cigarettes ran out and prices slumped still further; in February the supplies of food parcels were exhausted and the depression became a blizzard. Food, itself scarce, was almost given away in order to meet the non-monetary demand for cigarettes. Laundries ceased to operate, or worked for £s or Reichsmarks; food and cigarettes sold for fancy prices in £s, hitherto unheard of. The Restaurant was a memory and the BMk a joke. The Shop was empty and the Exchange increased in volume, becoming a larger proportion of a smaller volume of trade. This, the first serious and prolonged food shortage in the writer's experience, caused the price structure to change again, partly because German rations were not easily divisible. A margarine ration gradually sank in value until it exchanged directly for a treacle ration. Sugar slumped badly. Only bread retained its value. Several thousand cigarettes, the capital of the Shop, were distributed without any effect. A few fractional parcel and cigarette issues, such as one-sixth of a parcel and twelve cigarettes each, led to momentary price recoveries and feverish trade, especially when they coincided with good news from the Western Front, but the general position remained unaltered.

By April 1945 chaos had replaced order in the economic sphere: sales were difficult, prices lacked stability. Economics has been defined as the science of distributing limited means among unlimited and competing ends. On 12 April, with the arrival of elements of the 30th U.S. Infantry Division, the ushering in of an age of plenty demonstrated the hypothesis that with infinite means economic organisation and activity would be redundant, as every want could be satisfied without effort.

Human geography has long been concerned with explanatory descriptions of areal distribution patterns, often presented as distribution maps. The geographer attempts to explain such patterns resulting from the spatial organisation of man's activities, and in this way there is a bearing on the question of why men live in societies. Geographers have also long used sample or case studies to assist in their explanations, with the idea that an intimate view of the functioning of individual units throws light on the reasons for the pattern as a whole. From one viewpoint, such case studies lead into studies in human ecology, using ecology in a sense lying closer to the biologist's than the sociologist's usage. A complementary view, however, would link such case studies with economic flow patterns, market influences and the like.

It is in this context that we present the two complementary papers by Brookfield. The first is a general description of the terrain of the Chimbu people, and of the changes wrought during the last forty years in their society, economy and polity. Significantly, it was written as a contribution to a volume of collected papers called 'Geography as Human Ecology'. The analogy is with the biologist's ecology, in which he studies the web of interaction of plants and animals in a particular site; Brookfield's work in Melanesia over many years has shown how much depth can be imparted to the study of distribution patterns in human geography by an ecological approach.

'Geography as Human Ecology' was compiled for use in the last years of secondary and possibly the first year of tertiary education, while 'The Money that Grows on Trees' was first presented as Brookfield's presidential address to the Institute of Australian Geographers, a body consisting of professional geographers. Level and style are understandably different, but Brookfield also sets out to assess and experiment with the networks and systems approach to studying societies of relatively simple economy such as the Chimbu. This need not exclude cross-fertilisation with analytical studies of economically advanced societies.

THE CHIMBU: A HIGHLAND PEOPLE IN NEW GUINEA

H. C. Brookfield

INTRODUCTION

In the central highlands of New Guinea, living in a string of intermontane valleys that occupy a linear distance of over 600 miles from east to west, is a population of about a million people whose existence was almost unsuspected as recently as 1925. New Guinea is nearly 1300 miles in length, its backbone formed by one of the world's great ranges, a chain of fold mountains that reaches a maximum height of 16,500 ft in West Irian, where four peaks carry permanent ice. In Australian New Guinea the highest peak is Mt Wilhelm (14,600 ft) which, together with a number of others, although glaciated during the Pleistocene, now receives only rare and ephemeral snowfalls.

The people described in this essay live just south of Mt Wilhelm, occupying the sides of two steep, narrow valleys draining south from the mountain, and a wider more open longitudinal valley in the centre of the ranges (Fig. 4.1). These valleys are the home of the Chimbu people, one of the most numerous of the hundred or so language groups which make up the population of the central highlands. Of the 160,000 inhabitants of the administrative area known as the Chimbu Division of the Eastern Highlands District, some 60,000 are speakers of the main Chimbu dialect, and all but some 10,000 of the remainder speak dialects which are closely related. There are also a small number of Chimbu speakers in the Minj Subdistrict of the Western Highlands District, adjoining the Chimbu Division on the west. A distinctive feature of the Chimbu Division is the high density of population; it averages over 200 per square mile throughout the country below the forest line and rises locally to about 500 per square mile – a value paralleled only in one other part of the central highlands, the Baliem valley of West Irian.

The Chimbu were 'discovered' by an exploratory patrol

consisting of a government officer and a small group of gold prospectors in April 1933. Missionaries arrived soon after and, following the murder of two of these, a government post was set up late in 1934. By 1940 intertribal fighting had been halted, a rudimentary system of government through native officials established, a preliminary census taken, and the first school set up. The war led to some setbacks, but by 1966 there were five

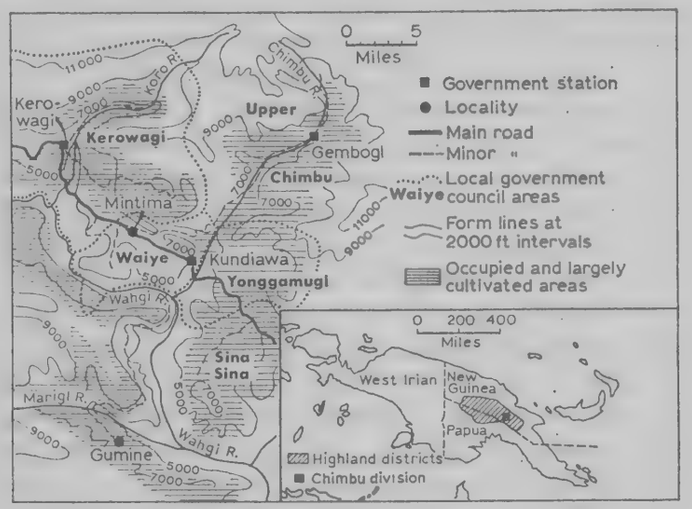


Fig. 4.1 Part of Chimbu Division, Eastern Highlands, New Guinea

government stations, rudimentary health services, several schools, a network of roads, and cash-crop agriculture had been firmly initiated. A few young Chimbus are now literate in English; a much larger number are fluent in Melanesian Pidgin, the *lingua franca* of the Territory, and have had some schooling. Several thousand men have travelled to the coast to work on plantations and in February 1964 the Division sent four elected members to represent its people in the Territorial House of Assembly in the capital, Port Moresby. In thirty years there have been great changes, and though the rate of progress is currently unsatisfactory, in that economic advance, health and education all lag behind other parts of the Territory, the transformations that have been and are being wrought in Chimbu society are revolutionary.

THE CHIMBU – THEN AND NOW

Today the people speaking the 'Chimbu' language think of themselves as one people, different from nearby peoples who speak different languages. This sense of unity is, however, entirely a product of the last generation; in 1933 there was no such feeling of common identity, nor any common name to describe either language or people. The very word 'Chimbu' is a name given by European explorers and seems properly to refer to a short stretch of the river, now known as the Chimbu, near the point where it was crossed by the first exploratory patrol.

Chimbu social structure is broadly speaking of the type known to anthropologists as segmentary; it is built up of a hierarchy of groups, each group of a higher order being subdivided into two or more segments of a lower order. Such groupings emerge in several ways. Most lower-order groups have some tradition of common descent in the male line; they also combine for certain purposes connected with work or ceremony, they have territories in common, they bear group names. Most significantly, however, groups of the same order of segmentation are recognisable as they emerge in opposition to one another in disputes or in certain ceremonies. Thus two *clans* may each be divided into four *subclans*. A dispute between members of two *subclans* of the same *clan* will merely mobilise support from among the subclansmen of the disputants. On the other hand a dispute between members of two *subclans* in different *clans* may, if it is weighty – or noisy – enough, involve all the members of the two clans, each supporting their own clansmen.

Anthropologists, like other social scientists, differ in their use of terminology, and there is no single accepted terminology for all the various orders of grouping encountered among the many societies of New Guinea. Paula Brown, an anthropologist who has worked in Chimbu with the writer, has advanced the terminology that is used here. Individual families are grouped into small local groups which she terms *subclans*, and a number of *subclans* together form a *clan*; in turn, two or more *clans* are combined in a *tribe*. There are also some more variable intermediate orders of grouping. Subclans number from as few as 30 to as many as 400 individuals; they occupy a common territory which they combine to defend and to enlarge, they carry out some work projects as a

group, they combine for ceremonial functions of all kinds and unite their resources at such functions whenever payments or presentations have to be made to other groups. They have an ideology of common descent in the male line but are unable to support this by any detailed genealogical record; indeed every subclan contains a number of men who were born as members of other subclans and who have been adopted into their present group, either in childhood or in adult life. Clans combine on fewer occasions than do their component subclans, but they are the main classificatory unit for purposes of marriage: the male and female natal members of any one clan may only marry with members of other clans. When clans grow large they sometimes split into two, some subclans forming one new clan, the remainder forming the other; marriages may then take place between members of the two new clans. Clans also have common territory, and combine to defend this territory against attack by members of other clans.

Tribes are territorial groupings of clans. They existed in the past principally as war-making units, but there are also certain ceremonies that take place as between tribes, and not between smaller units. These ceremonies, large and small, have a most important place in Chimbu life. They are the occasion of presentations of valued goods and pork between men who owe one another some mutual obligations; they are, in fact, the cement which binds the whole system together. Thus, when a girl marries, a payment is made by the subclan of the groom to the subclan of the bride; the payment is contributed by individuals and distributed to individuals. It is given not for the girl as such, but for the labour and expense of rearing her for the benefit of the groom's subclan. However, there are reciprocal benefits, for the man who marries the girl thereby gains access to the assistance of his in-laws, and this may take the form of hospitality, or a loan or gift of land, or of sanctuary in the event of defeat in war or some other personal disaster. These mutual obligations must be periodically recognised by the public presentation of valuables at a ceremony; the giving of Christmas and birthday presents is, in some measure, a parallel in our own society.

The Chimbu in 1933 were a Neolithic people, having no knowledge of metals and using tools of wood and stone. Their stone tools were axes, of sharpened polished metamorphic rock, traded from areas of the highlands possessing stone suitable for the manufacture of axe-blades. There were two types of axes – the heavy-bladed work-axes, used for chopping down trees, making fence posts and houses, and wielded in warfare, and the fine-bladed, nicely mounted ceremonial axes which were used only as valuables and which had no utilitarian function except occasionally in fighting. Wooden tools included spades, digging-sticks, spears and bows and arrows. With these tools the Chimbu and other highland peoples cleared the forest and tilled the land.

The principal crop was and is the sweet potato, covering the greater part of the tilled area, but a whole range of subsidiary crops was also grown, including especially sugar cane, bananas, a variety of green vegetables and some root crops like taro and yam. Before discovery, some maize was grown, and since discovery a whole range of new crops has been added to the system. The origin of the sweet potato in the New Guinea highlands is an unsolved problem. At the time of discovery it was the staple crop almost everywhere and had been so for many generations. Yet the balance of evidence suggests that the sweet potato was introduced into the Western Pacific from its American areas of origin only after the voyages of the European explorers in the sixteenth century. It must have reached the highlands through the tenuous trade routes across the mountains along which shells (used as valuables), the plumes of some birds and a few other goods entered in pre-contact times. If this is so it is hard to believe that it could have been established in the highlands longer than 300 years.

Long before 1933 the montane forests had been cleared entirely, except on a few steep slopes, up to heights of from 7000 to 9000 ft and had been replaced by a mixed landscape of cultivation and fallow. Cultivated areas, either the regular squares of sweet potatoes or the mixed gardens in which subsidiary crops were grown, were surrounded by fences to protect the crops from the pigs which, though fed nightly on sweet potatoes at the houses to which they returned to sleep, would wander freely by day among the grass of the unenclosed fallow areas. Most fallow areas were in grass, but there were also stands of slender casuarina trees, planted in old garden sites to provide firewood and timber for fences and houses and to aid recuperation of the soil.

There were no villages. The men lived in communal houses generally located on ridge-tops commanding a wide view, while the women shared, with the pigs, houses scattered along the fences

of the cultivated areas, and often buried in the foliage of tall sugar canes and trees. Here and there would be burned areas, fired by raiders, and a few more thoroughly devastated tracts from which the inhabitants had been driven in flight. No man went anywhere without his weapons, and raids and ambushes were of frequent occurrence.

Clothing was scanty, and never-washed bodies were smeared in pig grease for warmth against the cold of the mountains. Houses were low structures of wooden walls thatched with grass: at high altitudes they had double walls with an insulating layer of leaves of the pandanus palm. Life was harsh; perhaps half the children died before reaching the age of five and sudden death was an ever-present possibility. The glamour, the shouting, the finery and the excitement associated with warfare and with the occasional great intertribal ceremonies were the principal relief in a life of drudgery, danger and discomfort.

Warfare was suppressed in most parts of Chimbu by 1940, and though intergroup hostilities still occasionally lead to brawling and murder, the 'time belong fight' is now little more than a memory. A whole generation has now grown up without warfare. Clothing is now quite common, especially among the men, and a majority of the young people wash quite frequently. The carrying of weapons is becoming rare, more fallow land than formerly is now planted in casuarinas, far more land is cultivated and the diet is more varied. Steel tools have wholly replaced stone and, for diversion in the years before they settle down, young men now go to the coast to work as labourers, or seek out employment with Europeans as domestic servants, enlist in the police or, if they belong to the educated 'élite' with a few years of formal schooling, may join the army or follow a trade such as carpentry or truck-driving. In 1955 coffee was introduced to the Chimbu and has now become a major cash-crop, grown by almost every married man. But while there have been many changes, other aspects of life are little altered. Sweet potatoes still form more than four-fifths of the diet, most women still share houses with their pigs, most houses are still built low to the ground and are verminous and filthy, a high proportion of children still die in infancy and, for women, most of life is still a round of drudgery. The missions are widely attended and their counsel frequently obeyed, but there is little joy to replace the excitement of war; even the periodic ceremonies are slowly dying out. Though most Chimbu now have a little money, and some few have several hundred pounds, incomes are in general small and speedily eaten up by the demands of taxes (£2 a head in the central area), clothes, tools, utensils, the inflated payments necessary in marriages and fines for fighting or gambling at cards.

The landscape has changed greatly: isolated cultivated plots have been thrown together in great, continuous enclosed areas bounded by fences miles in length; large areas are under coffee, grown in the shade of casuarina trees; with little use made of fire, trees have everywhere extended their ground at the expense of grass; the housing of men especially is more scattered than before. The basic methods of cultivation, however, have not changed and all work is still done by hand, without aid from either draught animals or machinery. Fences are still built laboriously of sharpened stakes of casuarina wood, ditches are still dug by hand and garden produce is still carried home daily in bags, weighing perhaps 30 lb, suspended from the heads of the women. Thus the changes in Chimbu life that have taken place since 1933 have been only partial; while some aspects of life and of land use have been revolutionised, others have changed only little.

THE CHIMBU HABITAT

The central cordillera of New Guinea is a recent fold-structure which was uplifted in a series of stages culminating in the late Tertiary. There are no *nappes*, but imbricate folding is very common throughout the ranges. In the Chimbu area there are two parallel ge-anticlinal arches in the cores of which granites and granodiorites are exposed at the surface, forming two massive ranges – the Bismarck mountains, rising to 14,600 ft, to the north, and the Kubor range, rising to 14,000 ft, to the south. Between is a synclinorial trough merging westward into a fault-bounded valley occupied by the upper Wahgi river. The northern limb of this trough is complicated by severe strike-faulting so that the same formations are repeated over and over again. The southern limb, though also faulted, is more regular and also wider. Within this trough a succession of Lower Cretaceous to Miocene strata is preserved; tuffs, shales and greywackes predominate in the lower

beds, but a great thickness of very pure hard limestone occurs towards the top.

The Chimbu homeland is clearly divided into two parts – a mountainous north and a more open south. Northern Chimbu is composed of the two deep valleys of the upper Chimbu and Koro rivers, both draining from the Mt Wilhelm massif and lying in the deep gorges 3000 ft below the level of the high-flanking spurs. At the head of the Chimbu valley is an open basin developed entirely in granite with steep ridges and narrow valleys down which pour clear mountain torrents uniting at Gembogl to form the Chimbu River. The valley floor drops swiftly from 7500 ft at Toromambuno to 6500 ft at Gembogl. Below Gembogl the valley narrows and, down to the defile through the Porol range, traverses heavily faulted country passing first through metamorphic rocks, then through alternating bands of steeply inclined limestone, mudstone and shale. Side valleys occupy the less resistant formations, while limestone bluffs tower over the central gorge.

Immediately above the passage through the Porol range, two major side tributaries enter the Chimbu – the Kwinigl from the east and the Singganigl from the west. Forming the southern wall of these valleys is the most remarkable relief feature in all Chimbu – a stripped limestone backslope, angled between 35 and 45 degrees, cultivated throughout, that towers 2000 ft above the river.

This is the backslope of the Porol range, the crest of which divides the mountainous north of the Chimbu country from the more open south. Below the range lies a wide, deeply dissected vale which extends from the Wahgi valley to the Chimbu and beyond to the east. The crest of the Porol range lies at about 7400 ft. Its scarp face, with a slope of between 50 and 75 degrees, varies between 400 and 600 ft in height; below it to the south is a series of enormous arcuate landslips and then a tumbling footslope, characterised by great spurs and trenching valleys in alternating mudstones, shales and greywackes. The bottom of the vale lies below 5000 ft; in the hills rising beyond it to the south, the Wahgi has cut a gorge through a massive south-facing ridge of sandstones, greywackes and tuffs which lie against the slopes of the Kubor range.

The footslope of the Porol range is an area of active landslipping, and many of the upper slopes are overlain by block-flows or rubble-flows composed largely of limestone blocks and derived soil which locally extend more than a mile across the mudstones. These flows are of varying ages and, though none in Chimbu are evidently recent, similar flows have taken place within very recent years in other parts of the highlands. The flows are continually being resorted by minor landslipping and slumping.

Except in the west, where Chimbus have settled on the alluvial and colluvial terraces bordering the middle Wahgi, most of the Chimbu habitat lies in a zone of steep slopes. Streams have cut deep gorges, and even the channels of minor streams are deeply gouged into the landscape. Gently sloping country is rare, and occurs only around 5000 ft – the level of still-stand represented in the floor of the Wahgi valley and in other valleys in the highlands.

Chimbu lies 6 degrees south of the equator, and has a tropical highland climate with night temperatures falling regularly below 50°F above about 6000 ft. Even as low as 5000 ft night temperatures fall into the forties during occasional cold spells, and above about 8000 ft temperatures may fall below freezing-point in hollows that collect cold air. Farther west in the central cordillera, where the upland belt is wider than at Chimbu, freezing temperatures may be experienced at much lower altitudes, and even at 5000 ft air temperatures only a few degrees above freezing-point have been recorded. Normally at about 5000 ft in Chimbu the range of temperature lies between the middle fifties and the middle seventies on normal days, exceptionally failing to reach a maximum of 70°F and still more rarely rising above 85 degrees.

Rain falls on most days in a normal year, though periods occur with as long as twenty days without rain. Drought occurs most commonly between the months of June and November, and the months of February, March and April are almost always wet. Table 4.1 shows the mean monthly rainfall at five stations in the Chimbu area, four lying along a north–south section down the Chimbu and lower Wahgi rivers, and the fifth, at Kerowagi, lying farther west. Locations are indicated on Fig. 4.5.

The higher rainfall in the west is confirmed by other nearby stations. In general the Chimbu valley area, protected by high mountains on both sides, seems drier than the more open country around, though the differences are small. At all stations there is a well-marked peak in the mean figures in the early months of the year and June and July are the driest months, though in individual

years the driest period may occur much later. There is a slight tendency towards a secondary maximum in September–October. The climate is thus wet, with only a weak tendency towards a dry season and a liability to only intermittent short droughts. Cloud cover is heavy, and there are few days in the year when cloud on the surrounding mountains does not spread out over the intervening valleys during the afternoon. Many early mornings, however, are clear and cool, with inversion fog lying in all valleys; sometimes this ground fog persists until around 10 a.m.

mantled by the lush *Ischaemum* grass. Extensive fallow areas carry planted and spontaneous *Casuarina* trees.

Soils vary greatly from place to place according to both altitude and parent material. At higher altitudes, with lower temperatures, humus accumulates in the soil, and everywhere above about 6000 ft the organic matter content of the soil is higher than at lower altitudes. Some soils around 5000 ft have concretionary layers, which make them difficult to work. Most soils are formed on steep slopes which are subject to constant movement; they are therefore

TABLE 4.1

MEAN MONTHLY RAINFALL (IN IN.) AT CHIMBU STATIONS
(number of complete years in brackets after station)

	Jan	*Feb*	*Mar*	*April*	*May*	*June*	*July*	*Aug*	*Sept*	*Oct*	*Nov*	*Dec*	*Year*
1. Keglsugl (6)	9·82	12·22	11·17	10·65	5·29	3·81	4·05	5·39	6·11	6·05	5·79	9·61	89·96
2. Gembogl (2)	2·05	6·02	7·47	7·16	4·93	3·85	4·06	4·34	7·78	6·69	4·00	8·47	66·82
3. Kundiawa (7)	10·41	8·97	11·56	10·21	4·95	4·31	2·78	4·89	6·74	6·72	6·44	8·03	86·01
4. Gumine (3)	10·06	11·89	10·24	10·96	7·42	4·54	3·98	4·87	7·22	8·27	5·01	9·40	93·86
5. Kerowagi (9)	9·92	10·90	11·22	13·01	8·81	4·91	6·32	8·35	10·59	9·46	8·65	12·36	114·50

The wild vegetation is lower montane rain forest, dominated below about 7000 ft by species of oak (*Quercus* and *Castonopsis*), and above this altitude by southern beech (*Nothofagus*); some tall *Araucaria* pine is also found at the lower altitudes. There is one species of *Eucalyptus*; otherwise the vegetation has mainly Indo-Malayan affinities. At high altitudes, in areas covered by persistent, almost daily cloud either near or on the ground, is a belt of mossy forest, gnarled and heavily festooned in mosses. Above about 12,000 ft is an alpine tundra, and above 13,500 ft even this becomes sporadic. On most areas below 8000 ft, however, the wild vegetation has been replaced by gardens and fallow. The fallow vegetation is mainly grass, but of different species. At higher altitudes and in moist locations, a dense stand of plumed swordgrass (*Miscanthus* spp.), up to 8 ft tall, dominates almost everywhere. At lower altitudes and in drier, well-drained areas the dominant vegetation is one of short grasses, dominated by *Themeda* spp. In wet places reeds (*Phragmites* spp.) and the wild 'pit-pit' (*Saccharum spontanaeum*) form dense stands, while moist hollows are

permanently immature and reflect the dominant influence of the parent material – clay, sandstone, limestone, metamorphics and granite. Fertility ranges from moderate to low, *p*H from almost neutral to acid.

The Chimbu understand their soils well, and name them. Their system of nomenclature does not follow any modern system of classification but depends on texture and colour, or some combination of these. Thus friable, crumb-structured soils, usually dark in colour, such as occur on the limestone-derived landslips and block-flows, are known by the texture name *wīmen*; if they are very dark and rich in organic matter they are called black – *kāmā*. Clay soils are called *kogua* and pebbly soils *siugl*. Other soils are termed red, *bogl*. The Chimbu are ready to discuss the relative merits of their soils for growing crops and trees; they have general terms for deep fertile soils and for thin hungry soils. There also exist differences of opinion as to which soil is best for sweet potatoes; some prefer the light limestone soils for their ease of working and their ability to produce crops quickly, but others

prefer the clays which produce fatter tubers. Rocks, too, are well understood by the Chimbu, who can name all different rock types found in their area and describe their location in the surrounding country. There is good understanding of underground drainage in the limestone areas, though few will venture into the caves. There is good reason for this, for persons dead or dying of certain feared diseases, or those accused of sorcery, were thrown into such caves in the past, and their spirits are dangerous to all who are related in any way.

The feared diseases seem to be largely nutritional in origin, for the diet is very poor in protein; cancer is also feared. Fever, on the other hand, is not regarded as contagious. Malaria, the commonest fever, is seasonal in occurrence at this altitude and there is little or no transmission above about 6000 ft, where night temperatures become too low for the parasite to complete its life cycle. Today the association of malaria with mosquitoes is well understood; in former times the disease was believed to be due to spirits, and areas known to be liable to malaria infection were shunned, rightly but for the wrong reason.

AGRICULTURE

The Chimbu are subsistence cultivators, with a single-crop emphasis and little or no specialisation; their adjustment to environment is thus much simpler and more direct than is the case in an industrialised, commercialised society such as our own.

Chimbu married men and women spend the major part of their working hours in the fields. They also rear pigs, which are of importance particularly in interpersonal and intergroup exchanges, but in their call on both land and labour the pigs take second place to the crops.

There are two basic methods of land preparation in Chimbu. In the first, the whole land surface is turned over; in the second the land is trenched into a gridiron pattern of approximately 10-ft squares, and the spoil from the ditches is thrown on to the intervening beds. The former system is more common in the mountainous areas of northern Chimbu, the latter in the south and west. However, ditching is always employed on ill-drained land and in heavy soil, even on steep slopes, while complete tillage is commonest on friable, well-drained and easily worked land. The tool employed today is a steel shovel, but formerly either a sharp-pointed stick of casuarina wood about 6 ft long, or a spade made of hardwood, were employed.

Fire is often used as an initial aid to clearing, but only in grassland is it the normal practice to burn the whole area. Elsewhere, the use of fire is confined to the burning of heaped trash, especially branches and shrubs which are unsuitable for mulching. The initial work of land preparation, especially that of tillage or ditching, is carried out by the men. The men's work also includes the fencing of land to be cultivated against the inroads of pigs. Fences are made of sharpened stakes of casuarina wood, bound together by vines; they last from two to four years before they become weak and rotten. It is not now normal to fence each individual garden; rather we find large blocks of gardened land enclosed within a single fence (Fig. 4.2).

The fencing, clearing and digging of the land may be spread over a period of weeks, or even months. Finally, however, comes the time to plant the crop. The Chimbu make two types of garden, the single-crop *kongun*, always under sweet potatoes, with occasionally a little maize dibbled around the edges of the beds, and the mixed gardens, or *pongi*, which may have as many as twenty or thirty different crops, rarely including any sweet potatoes. The commonest crops in these gardens are sugar cane and bananas, but yams, cassava, taro, beans and green vegetables of various kinds are also grown along with some introduced European vegetables. There are also a variety of trees; some, such as the introduced pawpaw, are grown for the fruit while others are grown for their edible leaves. There are also two varieties of pandanus – an oil-bearing variety that is grown only below about 6000 ft, and the nut-bearing variety that is grown extensively above this altitude and also occurs wild in the high mountain forest.

The larger tubers from the sweet potatoes are used as the basic human diet while the smaller ones are fed to pigs. In terms of weight consumed, a pig probably eats about as much as a grown man; part of this diet is made up of grass browsed in the unenclosed areas, part of worms rooted from the ground, and the balance of sweet potatoes. The number of pigs varies greatly from one time to another; they are killed and consumed at all special occasions such as weddings and funerals and are slaughtered in

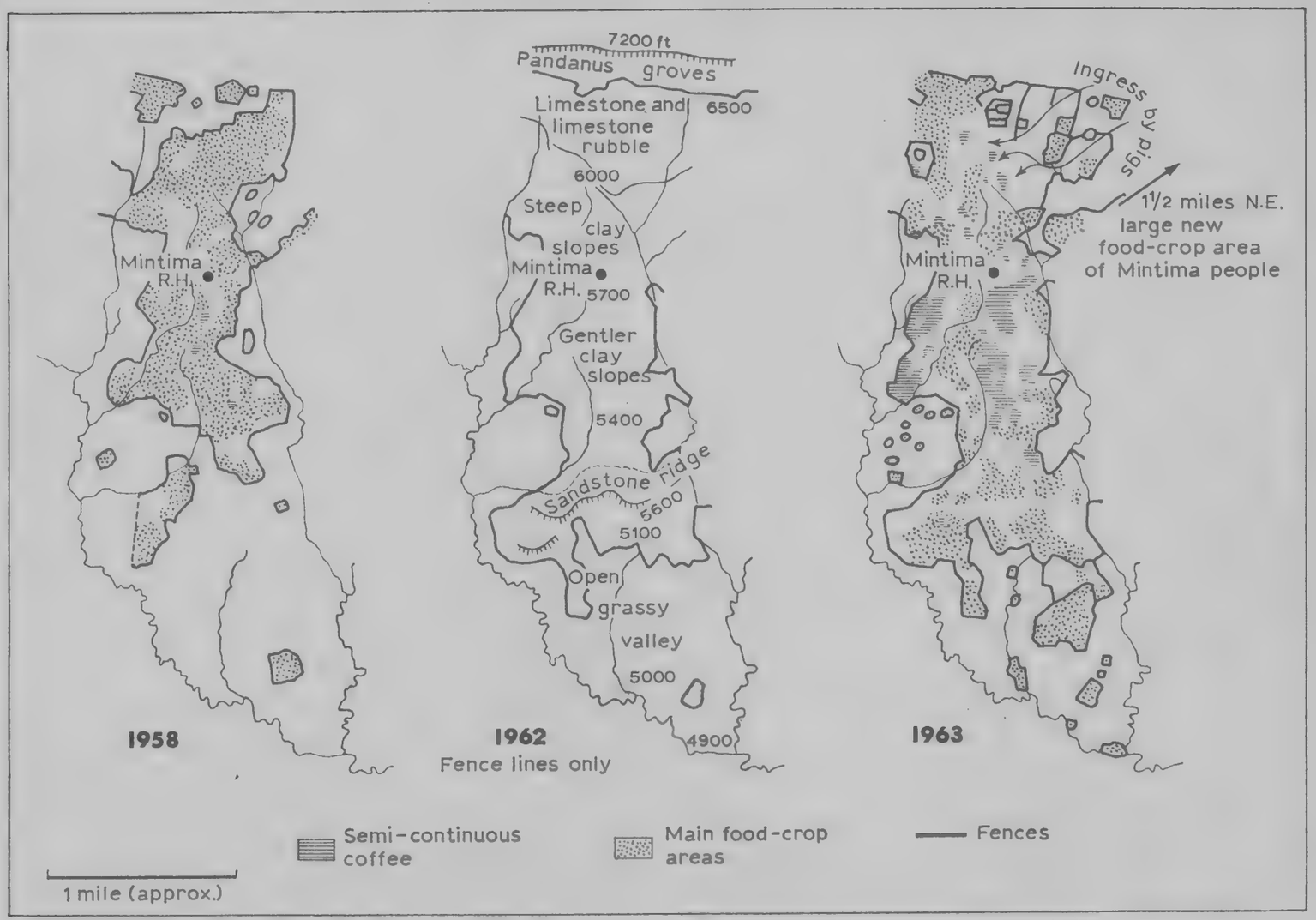


Fig. 4.2 The Mintima area, 1958–63

great numbers at the great intertribal presentations that are held every few years. The greatest numbers occur in the build-up period before these feasts; then there may be about 1·5 pigs per head of population and perhaps half the total production of sweet potatoes needs to be fed to the pigs. Thus the demand for this foodstuff fluctuates greatly from time to time, and while there is normally a surplus over requirements there are occasional periods of shortage which may be accentuated after a drought or a period of flooding rains.

Varying according to altitude and soil, sweet potatoes mature in from four to about ten months. As soon as a few tubers are fully grown they are lifted without disturbing the plant, and two, three or as many as five or six such harvests are taken at intervals until the plant is exhausted. The bearing period varies from as much as a year to as little as three months – the latter at high altitudes. Yields are difficult to determine, but vary between about 4 and 16 tons per acre per annum; the mode is probably around 10 tons. Since the area under all food crops in the mid-altitude areas is generally about a quarter of an acre per head per annum – perhaps 0·15 acres per head under sweet potatoes – production for human and pig use is thus around 30 cwt per head per annum, or around 10 lb per day in rough figures. This is probably a conservative estimate of present potential production, and it is clear that there is at most times an unconsumed reserve of this crop in the ground.

When a crop is finished, the land only occasionally reverts directly to fallow. Normally after the first two or three liftings of sweet potatoes little further effort is made to keep the garden weeded, so that in a few months it becomes choked by grass and other weeds. Pigs may be tethered on the plot to root up the remaining sweet potatoes, or, if the plot has been separately fenced, pigs may be admitted to roam freely around the garden and root up the whole area. Most commonly, the land is then reditched and replanted, sometimes with sweet potatoes, sometimes with mixed crops (*pongi*), though it is more usual for these latter to precede sweet potatoes on the land. This whole process may be repeated four, five or more times; numerous plots have been under semi-continuous cultivation since at least 1958, and it seems probable that there are areas of land that have known only short rest periods of two, three or four years at most during a time-span of two or three generations. For while Chimbu farming is land-rotational, in the sense that new gardens are constantly being made as old ones lie fallow, it is far removed from the simple shifting cultivation so common in tropical lowlands. There is much land that is cultivated for longer periods than it is fallowed and, over six years of observation, little evidence of a clear cycle has emerged.

When land is fallowed for a period of years, however, it is not always left simply to revert to natural growth. This is only done today in the thin-soiled areas dominated by short grass, and in certain wet and swampy tracts where a thick growth of tall reeds and canes springs up very quickly. On most land casuarina trees are planted with the last crop, and by the time this has been lifted the trees are firmly established and able to withstand the rooting of pigs. The casuarina is a quick-growing softwood which in favourable areas reaches a height of about 50 ft in less than ten years; it provides the wood for fencing, house building and fires, and it also restores fertility to the soil by adding humus and by fixing nitrogen by means of root nodules. As the Chimbu put it, the casuarina puts 'grease' back into the soil.

Casuarinas do well at all altitudes in the highlands except where the soil is thin and dries easily; on such areas these trees make very poor growth and are only rarely planted. These are the grassland areas, dominated in this part of the highlands by the waving clumps of *Themeda* or 'kangaroo' grass. They stand out clearly on air photographs.

The present distribution of enclosed and cultivated land in an area of central Chimbu is shown on Fig. 4.2, as are the changes that have taken place since 1958. In 1958 the traditional pattern of Chimbu farming was as yet not greatly modified, despite the replacement of wood and stone tools by steel axes and spades, and the small amount of production of sweet potatoes for sale to Government and Mission that had been in progress for some years. A number of new crops had been introduced and adopted and a small start made in cash-cropping with the introduction of passion fruit about 1953; however, this crop proved unmarketable and was soon abandoned.

In 1958 cultivated land in the area around Mintima Rest House was disposed mainly in a single fenced block lying between 6500 and 5400 ft; it occupied an area surfaced by limestone-charged

block-flows and landslips in the north, spread over steep clay slopes which form the surface formation farther south. Within the enclosed area rather less than half the land lay in fallow, mainly under planted casuarinas, while the balance was at various stages of cultivation. While there were several continuous blocks of gardens all at about the same stage, most of the cultivation and fallow were intermixed in higgledy-piggledy fashion. Outside the fences the pigs were free to roam, returning at night to houses located along the fences, with entrances on both sides, where they were fed a daily meal of sweet potatoes. Most pigs did not wander far from their houses, but browsed and rooted in the unenclosed area within a hundred yards or so of their own doors. Out in the unenclosed areas, however, was a thin scattering of isolated pig houses where pigs were also fed at night, roaming in the adjacent grasslands by day. Thus while limited tracts of the unenclosed land, especially in the southern area, were virtually unused, most of the land in the Mintima area was in some form of productive use, either for crops, for tree-bearing fallow or was grazed and rooted by pigs. There was clear soil selection in the distribution of land use; the most closely occupied and frequently used areas lay on the limestone rubble and the upper parts of the clay slopes, where soils are deep and well drained. The drier, thin-soiled areas of the southern valley were little used, while the deeply incised valleys of the south-flowing streams were not much cultivated because their steep sides are very liable to landslipping; but they were fairly closely occupied by the pigs.

Included in the cultivated area in 1958 were small patches of coffee, a crop first introduced in 1955. The acreage in this area, which has been kept under intermittent survey since 1958, has increased at an almost constant rate, and in the whole Waiye Council Area (Fig. 4.1) there were some 10,000 coffee trees by 1963. Since about 1959 coffee has been by far the principal source of cash income to the central Chimbu, though much is badly planted, still more badly maintained and almost all badly prepared for sale. By 1963 the whole pattern of agricultural distribution in this area, which is representative of a wide tract of country to east and west, had been transformed (Fig. 4.5).

By 1963 the size of the enclosed block had more than doubled, though it now included a far higher proportion of land in fallow than was the case in 1958. Much of the cultivated land on the deep-soiled clay slopes between 5500 and 6000 ft was now in coffee, the only significant exceptions being areas of new intake on the valley sides, and a tongue of thin-soiled land extending south-south-west from Mintima which has throughout borne little but sweet potatoes and grass. As a result of a build-up in the numbers of pigs and of the pre-emption of much of the former food-crop land for coffee, sweet potato cultivation had extended greatly and especially down on to the former grasslands in the southern valley, where large areas had been taken in for cultivation. Pig houses were still distributed along the fences and out in the unenclosed area, but far less land was remote from the pigs' regular ambit. At the same time the increased work demands of the new crop and the extended fence-lines led to some failure of upkeep of the fences guarding old cultivated areas. In 1963 a major breach occurred in the north-eastern corner, and large areas of sweet potato land were invaded and rooted up by the pigs. In 1964 this north-eastern quadrant was abandoned, and a new fence built farther back.

Present emphasis in central Chimbu is on the rationalisation of the coffee industry. Within the limitations of scant resources, the agricultural extension service is conducting a vigorous campaign designed to improve planting and maintenance. Central collection and drying points have been established at several places, including three within this small area, while a factory was established near Kundiawa by private interests in 1962 to prepare the beans for market and to grade them properly for sale. In 1964 this factory was acquired by a native company with a capital of £A37,000, organised by the co-operatives service of the Department of Trade and Industry.

But coffee is of value only to about half the 150,000 people of the Chimbu subdistrict. The 50,000 people living in the southern areas around Gumine are at present so remote from the lines of communication that they can find no buyers for their produce, while the 25,000 people having only land at high altitudes cannot grow coffee because their terrain is too cold. Recently pyrethrum has been introduced as a cash-crop for these latter, but the market is small and rather uncertain. European vegetables grow very well at these high altitudes but transportation facilities are poor and there is no large urban market in the Territory to create a demand. So a coffee belt (Fig. 4.5) is coming into existence

through central Chimbu and this belt is also the belt into which a money income is flowing; the outlying areas to north and south are not sharing in the general improvement. Whereas in 1950 there was virtually no difference in economic welfare between the different parts of Chimbu, by 1964 a regional differentiation, based on environmental suitability for coffee and degree of access to the main lines of communication, had already become well established.

POLITICAL GROUPS AND TERRITORIES

To write of political groups in Chimbu may seem surprising. Administratively the Chimbu occupy only one Division of one of the fifteen Districts into which Papua-New Guinea is divided, and their government is presided over by no higher a functionary than a District Officer. They send four members out of forty-four to the territorial House of Assembly. But in their pre-contact situation Chimbu had no central government at all, and the 60,000 people speaking the main Chimbu dialect were divided into some twenty-three 'political' groups – the *tribes* or alliances of *clans* – which have been defined and described in an earlier section. These *tribes* made war on one another and participated as a whole in major ceremonies, but they too had no central authority and at times there was fighting within them, even between small groups such as *subclans* with populations of only one to three hundred. The largest of the tribes numbered only some 5000 people. Yet in a sense we may consider the tribes as parallel to the states of our wider world, and the subdivisions of the tribe as parallel to the elements that make up a federal state.

In pre-contact times the frontiers between tribal territories consisted for the most part of wide belts of vacant no-man's-land within which it was unsafe for individuals to wander. Battles were fought with much shouting and display in these frontier zones, and sometimes defeat of one party would be followed by an invasion of their territory and perhaps by their wholesale expulsion. Over a period of perhaps forty years ending in about 1925 the Kamanegu dislodged the Naregu from the Singga valley (Fig. 4.3), driving them in a series of stages into their present habitat. Some clans and subclans went early; others remained until almost the end. Periods of warfare alternated with periods of truce in which exchanges might take place, marriages be arranged and some or all members of a defeated group be permitted to reoccupy some of their lost territory. Such truces might include the demarcation of a boundary line, where the frontier zone ran through valued and fertile land. Such a boundary might last for years; children would early be shown its position and warned that to transgress it meant almost certain death. From time to time young men might make raids across it, seizing pigs,

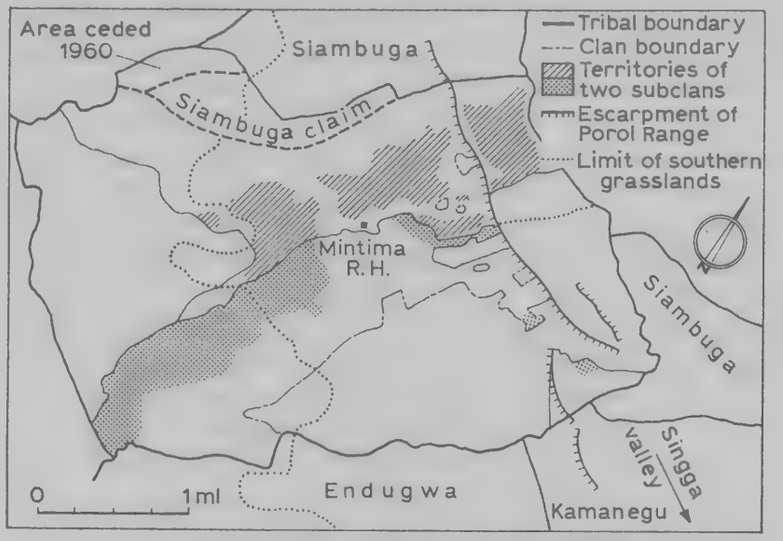


Fig. 4.3 Group territories within a Chimbu tribe – Naregu

violating women caught unprotected in the gardens, killing children, and perhaps attacking a whole community by stealth. Such raids might lead to reprisals, or escalate into war. Sneak attacks were not thought of as cowardly; men boasted of their achievements in terms of women and children killed in such raids. Life was accompanied by constant danger of sudden death. Open warfare was a matter of much bravado; battle would open by the shouting of insults and threats; men would assemble in full finery, fire arrows and throw spears and only a few would engage in hand-to-hand combat. Ambushes were the most deadly form of fighting; sometimes a whole party caught in this way would be massacred. When a tribal territory was deeply invaded, gardens would be destroyed, fences broken and pigs turned on to the sweet

potatoes, trees ring-barked and houses burned. Often whole populations had to flee, taking sanctuary with relatives in other tribes. This right of sanctuary was an important aspect of inter-personal relations. A multiplication of marriages between adjacent groups made warfare between them difficult, for men would not fight against men to whom they were related in marriage, and a marriage bond between men was also a basis for alliance. Thus some tribes would have many marriages with one neighbour, but very few or none with others of whom they were traditional enemies. The territorial pattern was in a constant state of ebb and flow, but there was a general movement of defeated groups out of the densely peopled Chimbu valley into the more open country to south and west. The eviction of the Naregu by the Kamanegu, described above, was part of this movement.

Within the tribe, fighting was less common and, generally, less vicious. Most marriages in any tribe took place between members of the different clans of the tribe, and there were a variety of bonds joining men together. Sometimes a single warrior would be skilful and powerful enough to unite a whole tribe under his leadership. Within the tribe, therefore, the territorial distribution of the various component groups arose as much from peaceful accommodation as from conflict, and the land available to a tribe is thus distributed among its component groups in such a way as to share out the available resources. Naregu tribe is composed of four clans of which one is only a recent addition to the tribe, being a remnant of a once much larger and more powerful group. The remaining three clans are divisible into a total of seventeen subclans, occupying a belt of territory extending from the Porol range down into the southern grassy valley; the Mintima area is a segment of this territory. The land may be divided into a series of belts, including the main slopes, the grassland at the bottom, a narrow shelf below the escarpment at the top, suitable for the growing of pandanus trees, and the forested face of the escarpment itself. The best land – the landslip and block-flow areas according to the Naregu's assessment – occupies only 12 per cent of the total territory, but every one of the seventeen subclans has a share in this land and also a part of the forest hanger on the escarpment above. Sixteen of the seventeen have access, either directly or through the unenclosed belts along the southward-flowing valleys, into the extensive pig-grazing areas of the southern valley. The remaining subclan has grazing land north of the escarpment, where fourteen of the seventeen subclans also have claims. Shares are not equal, but land is divided so that all have access to all the available types of land necessary to make up a complete farm (Fig. 4.3).

Territorial fragmentation is common within the tribe, where the need to defend holdings arose less frequently. Such fragmentation is most common on the most highly valued land. Land holdings in the Mintima area have been mapped in detail, and the size of these holdings has been measured. In the territory of one clan a block of enclosed or pandanus land includes 96 acres of land of the highest value and 210 acres of land of intermediate value. Taking the smallest segmentary groups – sections of subclans – as the territorial unit, the highest-value land was found to be divided into thirty-eight parcels with a median acreage of 1·00 while the land of intermediate value was divided into forty-five parcels with a median acreage of 1·55. One tract of limestone block-flow land, 150 acres in extent, contains holdings of members of nine subclans drawn from two clans. In the less valued areas on the lower ground, holdings are far larger and of simpler outline; boundaries are much less rigidly defined.

In recent years fighting as a method of settling disputes has been stamped out in Chimbu; though there are still occasional brawls, their immediate result is merely the filling of the local gaol. With this new *Pax Australiana* it has become possible to occupy former no-man's-land areas in safety and, both within the tribe and between tribes, belts of formerly vacant land have been rapidly occupied. An inevitable result has been an increase in land disputes and of the litigation that has taken the place of war. A Native Lands Commission was set up in the Territory some years ago to adjudicate in such disputes. In 1962 this became the Land Titles Commission with somewhat wider functions. Several Chimbu tribal frontiers have now been the subject of arbitration by the Lands Commissioner for the Eastern Highlands District, and other cases are still pending. One such dispute arose between the Naregu and the Siambuga, their western neighbours, in 1959–60, and the pattern of the dispute strikingly resembled a case study in international relations (Fig. 4.3). Each side marked out its claim and the Naregu were so incensed at the extent of their neighbours' claim that they pulled out the pegs, only to have

to replace them under the watchful eye of armed government police. Eventually the Commissioner arbitrated and awarded a slice of Naregu land to their neighbours. The decision was challenged with a show of force and only the presence of armed police kept the peace. In another dispute, between the Kamanegu and the Endugwa, feelings ran higher. The Kamanegu had driven the Endugwa from high country on the Porol range in the years before pacification but, following the peace, many Endugwa had returned into the lower reaches of the disputed area. They claimed restoration of the boundary of some fifty years ago, which they claimed to be the crest of the range. Eventually a decision was reached in favour of the *status quo*, thus denying almost all the Endugwa claim. A concrete peg was put in to mark the major turning-point in the boundary, but the same night this peg was uprooted; it was found next morning on the main road just outside the government station. It was replaced, and an armed police guard was mounted over it. Thus the decisions on which the peace rests are maintained only by a show of government force; though several tribes are now grouped into each of the Local Councils that have been formed since 1959, intertribal hostilities within the Council areas remain a powerful force. Though signs of its advent are not wanting, it will be some years yet before even so much as a Council area loyalty replaces mere tribal loyalty; it is easier for the Chimbu to give loyalty to wider and more amorphous groupings such as the whole Chimbu-speaking area, the whole of the Highlands, or the whole of Papua-New Guinea, than to political groupings that combine small and formerly hostile neighbouring tribes.

REGIONAL CONTRASTS

In pre-contact Chimbu there was little differentiation on economic grounds between the several parts of the area. All grew the same crops, and the only important differences were those arising from altitude, temperature and soil; the effect of these, on the very local scale of the Mintima area of Naregu tribe, has been discussed in some detail in the section on agriculture. On a wider scale, however, there were important differences arising from the density of population. In the upper Chimbu valley, mostly at altitudes between 6500 and 9000 ft, densities were and are very much higher than in the lower and more open country to the south and west. In some tribal territories today they rise as high as 450 per square mile, and in the more constricted territories before 1933, from which large frontier areas had to be excised as unoccupiable, effective densities may have been as high as 600 to the square mile or even more, even when allowance is made for the rather smaller total population of thirty-five years ago. In the lower and more open areas to south and west, few tribal territories have densities as high as 300 per square mile even today, and many have densities below 200 per square mile. There was thus much greater pressure on land in the upper valley and, in particular, there was no room for the extensive unenclosed pig-grazing areas that are found in the south and west. As a result an entirely different fencing system for the management of cultivation and fallow has been devised in the upper valley. Whereas, in the south and west, land may be divided into areas that are semi-permanently enclosed and areas that are entirely or almost entirely unenclosed, in the upper valley each quite small block of fields has its own fence, erected along lines of cordyline bushes that mark the permanent plot boundaries. Thus, when land goes into fallow, it is turned over to the pigs, and the unenclosed areas do not form a block as in the south, but a patchwork quilt of unenclosed tracts often linked together by narrow alleys between enclosed blocks. The approximate boundary between these fencing systems is shown on Fig. 4.4.

There were other differences. The rectangular gridiron ditching system of land preparation described above is found only in the south and west; in the upper Chimbu valley its use is restricted to quite small tracts of ill-drained land. Gridiron ditching is a practice found from Chimbu westward; eastward in the highlands and in northern Chimbu the practice is to turn over completely the whole surface of the land. Digging sticks were used formerly for this purpose rather than the wooden spades found in the south and west. This boundary too is shown on Fig. 4.4.

There are most significant differences between the settlement patterns of the Chimbu and those of their closely related neighbours to the south-east. All Chimbu-speaking areas have a dispersed settlement pattern consisting of scattered men's houses and dispersed women's houses. The individual woman's house is at a median distance of about 250 yds from that of her husband, but

in extreme cases this distance may be more than a mile. Though there is more clustering of houses in the densely peopled upper valley, the same dispersed settlement pattern, with a total absence of villages, holds true throughout. Immediately east in the Sina Sina area, however, and from thence eastward, population is clustered into large villages. A village might contain four or five

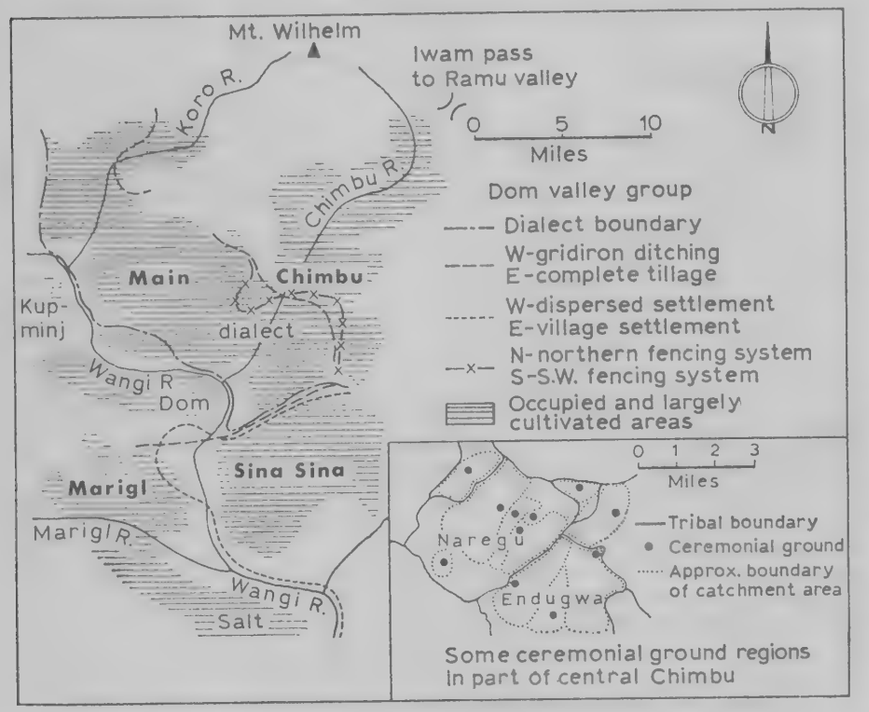


Fig. 4.4 Some cultural boundaries. *Inset:* Some ceremonial ground regions in part of central Chimbu

men's houses, each immediately surrounded by the houses of the attached women. Only pig houses are separate from the main complex. The boundary between these two forms of settlement is sharp, and it is interesting to note that those Chimbu who have migrated eastward into the next valley have adopted the village form of settlement practised by their new neighbours.

Contrasts in settlement pattern are accompanied by contrasts in ceremonial practice. The most important ceremonial events in Chimbu are the great pig ceremonies held every few years, when thousands of pigs are slaughtered and their cooked meat distributed among relatives and creditors over a wide area. These occasions are preceded by months of preliminary dancing and singing, leading up to the ceremonial climax in which mock attacks are made on the ceremonial ground and various fertility rites were formerly carried out, designed to ensure a bountiful supply of people, pigs and sweet potatoes alike. In the village areas east of the Chimbu river the ceremonies are held in the villages; in the dispersed settlement areas, on the other hand, the ceremonies are carried on in special grounds on which impermanent villages are built and occupied for the duration of the ceremony. These ceremonial grounds are focal points, drawing a population often cross-cutting clans. They vary greatly in the size of their temporary population, from as many as 1300 to as few as 80. The grounds and the populated extent of their catchment areas in a part of Chimbu are shown on Fig. 4.4.

Such trade as occurred in pre-contact central Chimbu largely took place in institutionalised exchanges and presentations such as those made at the great pig ceremonies. There was much more limited private trading, and a few itinerant hawkers from groups outside the highlands were able to travel quite widely – enjoying immunity from the local feuds and murder – trading in shells, plumes and other valuables produced outside the highlands. Stone axe blades, manufactured where special stone was available, were similarly traded within the highlands. However, on the fringes of the highland area there were certain recognised market sites where the surplus products of the highlands – pigs and stone axe blades – were traded at occasional markets for lowland supplies. Shells and plumes were the main imports. One such site was at the head of the Chimbu river, close to the foot of the track leading over the Iwam pass to the Ramu valley; its hinterland included most of Chimbu and areas to east and west. The stone axes traded here came from many miles to the west in the Jimmi valley, though axes from this valley also reached the lowlands from trading places situated north of Mt Hagen.

In post-contact times this trade has altogether ceased, and a new pattern of trade areas has sprung up focused upon government stations, missions, medical aid posts, schools and Council centres. There is also trade along the roads; motorised buyers purchase firewood, sweet potatoes and vegetables, and especially coffee,

from a number of recognised buying points. However, a high proportion of the goods sold off the land, especially vegetables, are still carried into the government station or mission for sale; between Kundiawa and the large Catholic mission at Mingende is a 'break point'; it is situated about twice as far from Kundiawa as from Mingende – a reflection of the relative buying power of these two places. Kundiawa, much the larger of the two, commands a wider and more populous hinterland. The approximate pattern of buying areas for vegetables in a part of Chimbu is shown on Fig. 4.5. It will be seen that quite large and populous areas are too remote from roads or centres to participate in this trade, while high-altitude areas cannot, for climatic reasons, participate in the coffee trade.

The trading foci that develop in present-day central New Guinea are the major airstrips, for, while the road carries some heavy traffic, it is unsuitable to serve as a major artery of commerce. Most of the road is hand-built and maintained, grades are steep, surfaces extremely rough or soft and the road is narrow with numerous sharp bends. All the by-roads are open only to four-wheel-drive vehicles after rain, and wheel-chains are often essential. The bulk of goods traffic in and out of the highlands still goes by air, and Kundiawa airstrip is linked by frequent scheduled and charter planes to both Goroka and Madang. Since only small aircraft can be used, costs are high by this method, and are much higher at the minor airstrips, which have severe load and other operating restrictions. Two of the Chimbu airstrips are remarkable; Keglsugl, at the head of the Chimbu valley, is the highest in New Guinea, while Omkalai, above the Marigl river, opposite Gumine, is the steepest. This strip has a gradient of 17 degrees, and a fall of over 300 ft from top to bottom. Aircraft use full power immediately on landing to keep moving up the strip, and on take-off continue to descend for a time after becoming airborne. Yet these strips are the central points of the whole communications system, serving as the main points for the import of consumer goods and the export of coffee.

In a wider sense, however, Chimbu lies in the hinterland of Goroka and Madang, and Goroka in turn lies in the hinterlands of Madang and Lae. By about 1967 there will be a two-lane all-weather road from Lae to Goroka, and at present a proposal is under discussion to build a second main access road into the highlands from Madang to Mt Hagen (Fig. 4.5). Chimbu will then lie at the limit of access by either route – perhaps an unfortunate situation for an area which is the most populous in the whole Territory.

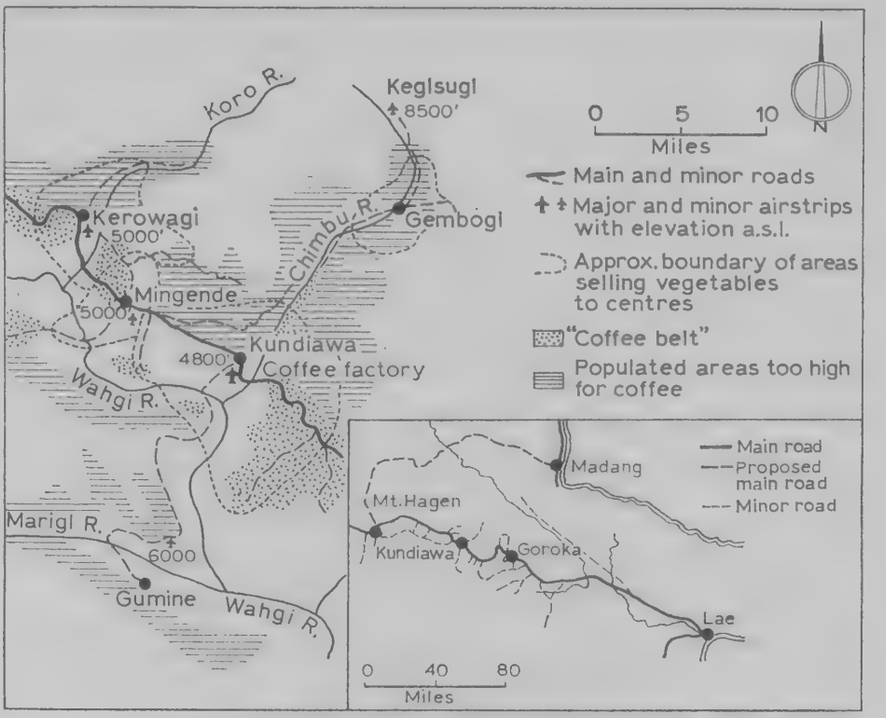


Fig. 4.5 Some modern regions

CONCLUSION

It would be wrong to generalise from this account of one small population, totalling only 5 per cent of the whole population of the New Guinea highlands, even about the neighbouring peoples to east and west. A research student working under the writer's supervision has carried out a study of three communities in an adjacent valley to the east; all three are different, and each is different again from the Chimbu. Their social structure and settlement pattern show differences, their land use has only certain basic similarities, and their system of land allocation shows important contrasts. Anthropologists have shown us repeatedly that the different societies in the New Guinea highlands are widely

contrasted in many respects, and there are even significant differences between different parts of Chimbu.

However, there are some elements that all societies and all environments in the highlands have in common. If we compare the detailed results of research carried out by geographers or anthropologists in different parts of the region, it becomes evident that the points of difference are narrowing down to quite small fields of subject-matter, and thus we can, by this method, build up a synthesis over a wider area. Furthermore, many of the same characteristics are found in other tropical areas outside New Guinea; comparison of results from detailed work can illuminate geographical inquiries in regions far apart.

Comparisons with the past are more dangerous but, given adequate precautions, may be made. The Chimbu were a stateless stone-age people until very recently; now they are an iron-age people subjected to strong central government and are in process of transition from a subsistence to a monetary economy. It will be obvious from Fig. 4.3 that there are comparisons to be made between the system of land allocation in use in Chimbu and the well-known 'strip parishes' of eastern England or northern France, or of the Québecois of the lower St Lawrence. Given an understanding of differences in culture, one can derive some interesting thoughts about the principles of land allocation in subsistence societies, operating to share out the available resources so that all have access to the land they need to make up complete holdings. Comparisons can also be made between the changes that are emerging from the introduction of a cash economy in Chimbu since 1955 and those which arose from its earlier introduction in, say, West Africa, Colombia, or that which followed the introduction of government, roads and towns in ancient Europe.

The geography of the Chimbu is thus presented here as an illustration of a principle: that by detailed studies of small areas, taking account of their culture and society, a deeper insight can be gained into all sorts of geographical relationships, and hence lead to enrichment of geography as a whole.

Note: A fuller study is H. C. Brookfield and Paula Brown, *Struggle for land: agriculture and group territories among the Chimbu of the New Guinea Highlands* (Melbourne, Oxford University Press, 1963).

THE MONEY THAT GROWS ON TREES: THE CONSEQUENCES OF AN INNOVATION WITHIN A MAN–ENVIRONMENT SYSTEM

H. C. Brookfield

From a paean of praise to the innovator:
When you first brought these trees and told us we should grow them, we scoffed at you and said 'We have plenty of trees already: what use are these new ones?' But now we see we were wrong: money grows on those trees.

A Chimbu orator, 1959, translated.

On the indivisibility of systems:
No one part of the model-system is more important than any other part. We may consider each model as a vital link in a communications system. We may identify feedback effects. Every change in technology or change in aspiration levels will lead to a reappraisal of the environment. Every movement of knowledge over space will need to be checked against environment and environment itself will change as a result of the decisions made. Changes in social structure may lead to changes in organisational behaviour and so on.

D. Harvey

PROTAGONISTS for and against the utility of general systems theory in geography are lining up thick and fast these days. Stoddart and Chisholm seem to polarise opposing views that have developed since Chorley threw down the first major statement on the subject, and since Ackerman offered the concept of the man–environment system as geography's overriding problem. Most writers quote or cite one or other of the numerous works on general systems theory by von Bertalanffy – my own preference would be for a late and concise statement of his – and efforts are made to show how systems, or ecological method in general, have a fairly long antecedence in empirical research in geography before they were specifically discovered *per se*.[1] But in all this there has been too little attempt to offer actual empirical evidence on the utility of any sort of systems approach in our more mundane tasks, evidence from which it might emerge that the particular

sort of system with which we deal calls for a methodology of its own that is neither a straightforward adaptation of ecological systems analysis, nor simple common sense, as Chisholm would have us believe.

Used as a tool of interpretation in human geography, systems are an order imposed on diversity in the real world; they are models of reality, and changing systems are models of changing reality. Here I seek to generalise about an event in the development of an underdeveloped country, involving a complex series of interconnected happenings through a period of time. I want to analyse the manner in which an innovation is not merely diffused in space, but also diffused up and down through all aspects of society and environment. To do this I make use of a simplified and partial system in which rising demand for goods of external origin is the source of energy. Model and reality are somewhat intermingled in what follows, but throughout there is a measure of simplification from the complexity of real events.

The event which I am discussing took place in the Chimbu district of Australian New Guinea after 1954, and I describe it from field observation extending intermittently from 1958 to 1967. The study area and population are tiny; almost all data are drawn from an area of 8·90 sq. km lying along the boundary between two exogamous clans within the territory of Naregu tribe, providing the main support of some 900 people (Fig. 4.6). The innovation has spread to cover only 66 ha. within this area by 1967. Very similar events were taking place in a much wider area, but I make only the most limited reference to this larger field. Since I have already written much about this area and its wider environment, I begin not with a description of environment, economy and society, but by merely outlining some elements of a hypothetical man–environment system that, broadly speaking, might stand comparison with Chimbu around the turn of the present century.

A pre-colonial system

The area is a tropical highland, a belted terrain of high relief, climatically humid with a rather variable tendency to a drier season in the middle and later months of the year. Environmental variety is well perceived and evaluated, and is reflected in a multitude of ways in the pattern of occupance. The area has been inhabited a very long time, but among a mobile population that fights over land, the present inhabitants of the study area have only recently moved in, displacing former occupants who have been driven to the west. Technology is Neolithic in terms of its tool-kit, but elaborate in its use of soil, so as to derive adequate subsistence with a 'normal surplus' in Allan's sense, under high population density that demands prolonged use and frequent re-use of the same land. In the short term, cultivation methods are well adjusted to the rate of recovery of soil nutrients; livestock are reared in large numbers and are periodically reduced by slaughter in major feasts. These feasts serve the purpose of recognising interpersonal obligations that cross-cut the segmentary quasi-descent groups into which the population is divided, and also of transferring goods of various kinds from one person to another. There is a small trade of surplus produce against external goods and valuables; new crops and new goods are periodically introduced and old crops disappear. Population is dispersed, though men are to some degree nucleated in communal houses; houses are built, decay and are replaced on the average about each four years; they are not always rebuilt in the same place, but the settlement pattern as distinct from location undergoes little major change. All this has been going on without revolutionary change for many generations.

THE INNOVATION PATH

Preconditions of the major innovation

In summary form, the flow of events is systemised in Fig. 4.7, where the events of the main innovation path and some concurrent events are set out. Pacification in the 1930s had effects throughout the system. The end of warfare made possible the occupation of frontiers, and thus released land; stone tools gave way to steel, and this change released some labour. Activity in the ceremonial and exchange sectors increased with greater freedom of movement. Velocity of circulation of scarce goods used in exchange thus increased, creating a demand for additional goods. Money was quickly adopted as a means of acquiring goods, but with the increase in competitive display traditional values continued to inflate, creating a demand for money in excess of supply. Sale of cash-crops, and recruitment for work first within, then,

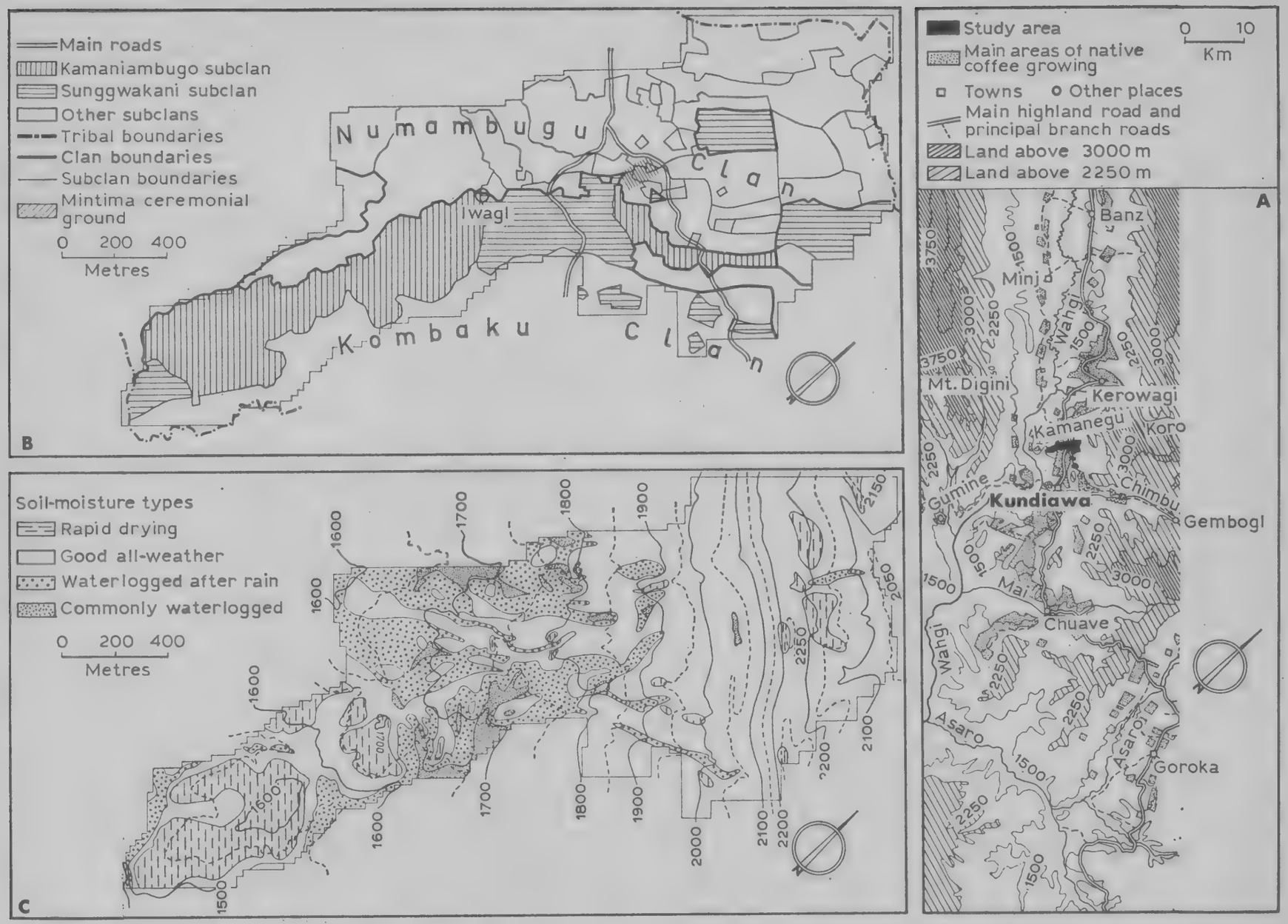


Fig. 4.6 (*a*) The central highlands of New Guinea and the Chimbu area
(*b*) Territorial divisions within the study area
(*c*) Altitude and soil-moisture conditions

Source: (*a*) is based on U.S. Army Map Service 1:250,000, sheets SB 55–5 and SB 55–9. (*b*) and (*c*) are based on field survey.

after 1950, outside the highlands, failed to satisfy the demand. This rising demand level, far more than the availability of land and labour, provides the essential preconditions for acceptance of innovations, and has continued as a basic condition throughout this story.

Introduction and ecological adaptation of coffee

All coffee grown in the New Guinea highlands is derived either from the agricultural experiment station at Aiyura, or from the Lutheran mission at Asaloka, near Goroka, places where it was introduced before World War II. All is *Coffea arabica*, mostly of the Typica and Bourbon varieties, the latter predominating in later diffusions. Coffee has a wide range of environmental toleration, but this range is somewhat restricted in the absence of fertiliser and skilled supervision. Arabica coffee is deep-rooting, intolerant of drought and waterlogging, of cold and great heat; slight acidity and good humus content are the most important soil conditions. Moisture requirements are comparable with those of the sweet potato, the existing staple crop in Chimbu, but higher yields are obtained on rather moister soils if these are well drained. Chimbu soon found that most coffee planted above 2050 m and on lower and drier country with thin hungry soils was a complete loss; initial failure to plant adequate shade also led to very heavy losses from die-back. Chimbu agricultural techniques excel in field drainage, and employ *Casuarina* as fallow cover; using these techniques it was quickly possible to use moist deep soils on shales disturbed by landslips, soils in which *Casuarina* made fast growth. Table 4.2 shows that these considerations were already substantially perceived by 1958, but that there has subsequently been more expansion on the moister than on the drier soils.

The first coffee was introduced by a policeman on leave, who stole it from Aiyura, about 1952. During the wet season of 1953–4 a selected group of leading men was encouraged by Government to plant coffee; one, living about two miles east of this study area, did so very extensively and became the principal innovator. By his drive and political skill he built up a following who, in 1955–6, set out four 'communal' blocks on land extracted from private holding, and planted nearly 4000 trees; the 'communal' blocks were quickly subdivided, and other individuals planted coffee on their own land. In 1957 a number of nurseries was established, and

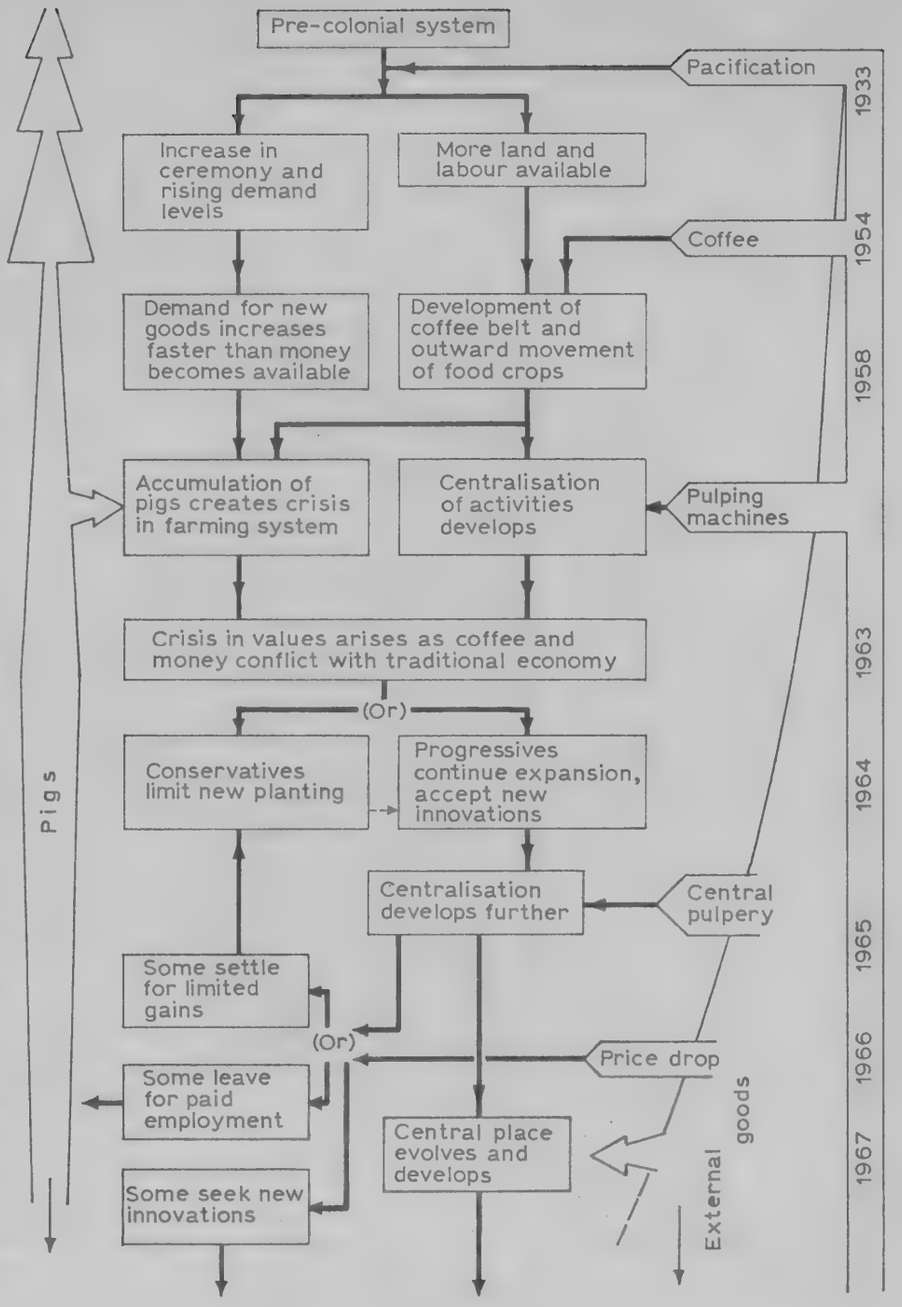


Fig. 4.7 A simplified model of the innovation path and selected associated changes

seedlings were widely planted out. Government patrols thinned planted coffee in later years, and the cut-down trees and other spontaneous growth provided adequate seedlings for all new planting after 1960. Much early coffee failed, but by 1961 almost all adult men had some coffee, and a few leading men had as much as a hectare. The rate of expansion is shown in Table 4.3, where it is clear that the early 1960s represent a peak period of new planting.

TABLE 4.2

ECOLOGICAL ADAPTATION OF COFFEE IN THE STUDY AREA

Percentage of coffee in each altitude class:	*1958* %	*1962* %	*1965* %	*1967* %
< 1600 m	6	1	2	2
1600–1752 m	62	46	60	60
1753–1904 m	26	40	34	31
> 1905 m	6	13	4	7
Percentage of coffee in each soil-moisture class:				
Rapidly-drying	3 } 48	0 } 41	1 } 41	1 } 38
Good all-weather	45	41	33	37
Waterlogged after rain*	52 } 52	51 } 59	59 } 66	57 } 62
Normally waterlogged*	0	8	7	5

*Unless drained

Source: From field survey, plotted or replotted on 1964 base at 1:10,900, gridded within 4-mm squares (= 0·19 ha.). Altitudes from a contract survey adjusted to four tellurometer determinations, and originally in feet. Soil-moisture from field survey at all seasons.

Adjustment and conflict within the farming system

Figs. 4.8 and 4.9 show the expansion of coffee in relation to food-crop land and to the variable position of the fences that separate enclosed land from the unenclosed areas in which pigs wander freely by day. Individual land holdings are scattered through group territory, and often include some borrowed land within and outside the territory of each individual's subclan. Coffee was at first planted on land that had borne food crops for some years and was ready to pass into fallow, but before long land was taken out of current food use for coffee so that new food-crop land had to be prepared farther out; large areas of dry grassland at lower altitude were enclosed for this purpose after 1959. Attention to coffee absorbed time as well as land, and a residential spread that would otherwise have followed expansion of food-crop area did not take place; men especially remained in the central area, close to most of their coffee. In the dispersed settlement pattern of Chimbu, each family farm has more than one central point – the separated houses of husband and wife. Since coffee was generally processed at the man's house, this became the centre of the coffee farm, and wife and pigs moved farther out. Table 4.4 shows the manner in which farm operations became ordered by distance.

TABLE 4.3

THE EXPANSION OF COFFEE-PLANTING

	Kamaniam-bugo (ha.)	*Sunggwak-ani* (ha.)	*Other groups* (ha.)	*Total study area* (ha.)
1958	2·3	2·3	3·3	3·9
1962	7·4	6·5	16·6	30·5
1965	11·8	13·5	33·5	58·8
1967	13·3	16·3	36·2	65·8
Mean annual increase during each period:				
1958–62	1·3	1·1	3·3	5·7
1962–5	2·2	3·5	8·5	14·2
1965–7	0·8	1·4	1·3	3·5

Source: From field survey as for Table 4.2. The areas measured include some coffee in outlying territorial blocks not included on the maps accompanying the paper.

Coffee yields were very variable, but on the basis of some data gathered in 1965, and with some large assumptions, it seems that the mean yield is about 0·7 tonnes per ha. of dry factory beans, with a range between 0·55 and 0·9 tonnes per ha. This compares with plantation averages around 1·1–1·2 tonnes per ha. in the Eastern and Western Highlands Districts. The best yields were obtained only by good planting, good maintenance and unusual expenditure of effort. Bad planting, poor maintenance, dispersal of holdings and very variable willingness to apply the work necessary to utilise fully all available ripe coffee, however, probably keep the mode to about one-half the plantation yield.

While coffee was becoming established, however, certain pre-

coffee trends were simultaneously in progress. Naregu tribe slaughtered most of its pigs in a major ceremony at the end of 1956, and during 1958–62 numbers of pigs were again increasing, demanding a growing share of the cultivated sweet potatoes, more foraging land and stronger fences for the protection of the crops. Progressive reduction in the area under special foods used largely in prestations, especially sugar cane and bananas, was an early symptom of stress, and by 1963 there was something of a crisis when pigs broke the fences at a number of points and spoiled

TABLE 4.4

CUMULATIVE PERCENTAGE OF CULTIVATED AREA DISTANT FROM HOUSE OF MALE AND FEMALE OPERATORS FOR A SAMPLE OF 26 FAMILIES

Distance from house (m)	*1958 All crops*		*1965 All crops*		*1965 Food crops*		*1965 Coffee*	
	M %	F %	M %	F %	M %	F %	M %	F %
0– 90	22	31	12	22	9	20	21	27
0– 456	72	56	55	49	48	47	76	54
0–1371	86	80	78	75	75	75	87	77
0–4500	100	100	100	100	100	100	100	100

Source: Modified from work by Brown and Brookfield. Straight-line measurements made on a map scale of 1 : 3600. Original measurements in yards.

large cultivated areas. Adequate new fences were not at once raised, but a temporary solution to the problems posed by the competing demands of coffee and pigs was provided by the creation of deep re-entrants into the enclosed area as a means of integrating the several sectors of the family farm more closely. But this created still longer fence lines, and the crisis in farm organisation merged into a wider reappraisal of their system in which the Chimbu became involved.

The conflict of values: 1963

By about 1963 the Naregu should again have been ready to organise a pig ceremony, and a strong body of opinion held that this should be done. The innovators, most of the Councillors, and others with the smell of money keenly in their nostrils were opposed and urged that all energies should continue to be bent towards the search for wealth. A minor incident flared into a fight following which many men, mostly traditionalists, were gaoled. Resentment against the progressives boiled after this incident, and first formal steps towards the organisation of a new ceremony were defiantly taken.

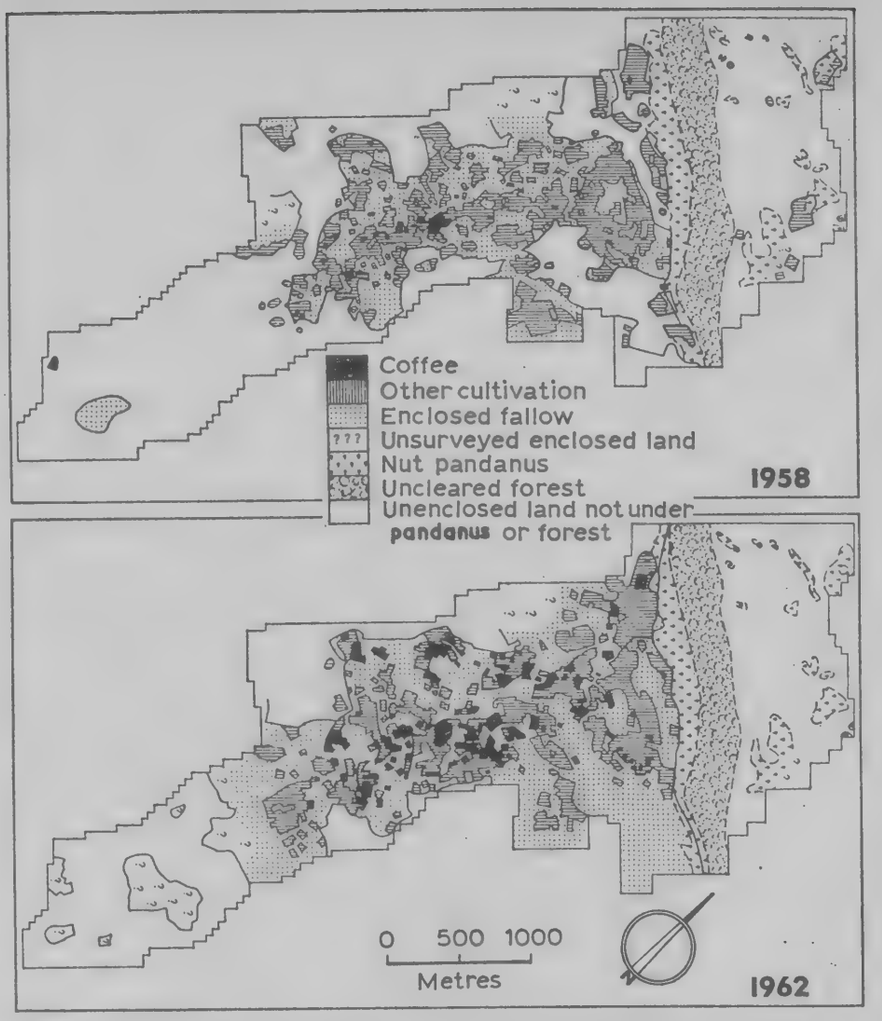


Fig. 4.8 Land use in 1958 and 1962 (from field survey, replotted on 1964 base)

Income from coffee was, however, becoming substantial, with keen competition among European buyers from the areas both east and west to purchase Chimbu coffee. The rising income was

still being absorbed largely in an inflation of traditional payments and prestations, but even a limited degree of European enterprise in opening new trade stores was generating greatly increased spending on clothes, utensils and increasingly also on foodstuffs. The traditionalists mostly found that a higher level of demand required them to continue to plant coffee, and most found their activities recycled back into the progressive path (Fig. 4.7). In consequence there was no large increase in the number of pigs

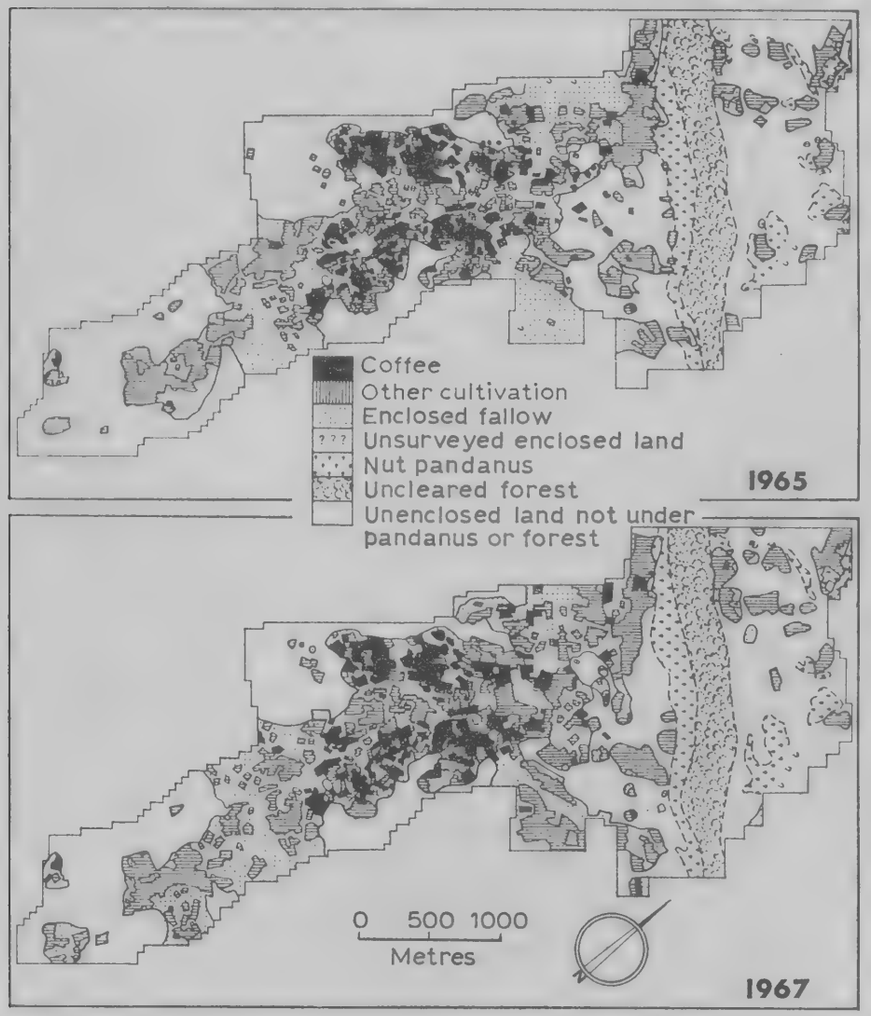


Fig. 4.9 Land use in 1965 and 1967 (1965 from field survey, 1967 from air photographs with partial ground check)

after 1963, and the traditionalist movement degenerated into a growing differentiation between men for whom coffee became a major enterprise, and men for whom it remained a minor but necessary adjunct to a mainly subsistence economy. For these latter coffee came to be regarded essentially as a bank, to be reaped and processed as and when money was required, but often neglected for long periods.

The victory of the progressives did not at once become apparent nor lead to any major reconstruction of the farming system. However, the regrouping of settlement into the central area now became more marked, and the earlier dispersal of farming activities was also replaced by concentration towards the centre; the area under special foods used in prestations diminished further. More fundamentally, purchase of imported food increased sharply: the then major trade store in Kundiawa reported a sharp upturn in sales of rice from about half a ton a month before 1963 to reach five tons a month by 1965. The victory was sealed by two new innovations in 1964. In that year the Department of Co-operatives sponsored a native-owned Kundiawa Coffee Society which bought a coffee factory at Kundiawa from its European operator, leading to a concentration of regional buying at a single point which quickly demonstrated that the Chimbu area was producing 12 per cent of all the coffee grown in New Guinea. Then there were also local innovations in the processing field, and these had important consequences for the spatial organisation of activity in the study area.

Centralisation of coffee processing: 1964–5

Coffee is processed by the wet system in New Guinea. The ripe cherries picked from the tree are first opened, or pulped to obtain the bean. This bean is enclosed in a viscous mucilage which has to be removed by fermentation. Pulping and fermentation must be begun within twelve hours of picking; the ferment in Chimbu generally lasts about 24–30 hours. After fermentation the beans are washed, and unripe or overripe beans float to the surface and are skimmed off. The beans are then sun-dried to a moisture content of about 12 per cent, a stage at which the bean can just be dented with the teeth. This is parchment coffee, which still has attached a yellow husk and an inner silverskin; these are removed by polishing which, together with final uniform drying and

grading, is a factory operation. For the native producer there are two stages: in the first the cherry is pulped to yield unfermented wet beans; in the second the beans are fermented, washed and sun-dried to yield parchment coffee which is sold to factory buyers. The weight proportions of cherry, wet bean and parchment are roughly 5·5:4:1, or alternatively 100:70:18.

Initially all coffee was processed at the men's house of the operator. Cherries were pulped with the fingernails in hollow drums in which the wet beans were also fermented and washed; beans were then dried to parchment stage on mats, or later on raised racks, then carried in sacks or other containers to buying points situated each few hundred yards on the main road where, after 1959, stands were set up for spring-balances carried by the buyers. Pulping by fingernail was a time-consuming job, so after 1960 some men, usually in local groups, bought hand-operated pulping machines of varying sizes, costing between $40 and $90. They charged outsiders for the use of these machines. The pulping machines were set up either near the men's house, or at a suitable location among the coffee. They required water for lubrication and to remove the pulp, and this was generally carried in bamboo tubes and applied in quite insufficient quantity. Fermentation, washing and drying were done in drums and on mats at the men's house, as before.

In 1964 the first central pulpery in the general district was set up in this study area, at Iwagl, with advice from the Agricultural Officer, and using funds collected by contribution and borrowing among two subclans – Sunggwakani and Kamaniambugo. At central pulperies cherry brought by individuals is weighed and recorded, then put through a pulping machine into a concrete fermentation vat. Running water carries the wet beans into the vat, and then out into lower washing tanks or flumes after the ferment. Finally, the washed beans are carried outside by water, collected, and spread on long racks which are supplied with polythene covers that can be drawn across at night or in rain. The uniformly dried parchment is then stored and sold in bulk. Payment is made from the proceeds on the basis of the proportion of cherry supplied by each producer. A clerk of sorts, and two regular employees, are needed for full-scale operation, and these must also be paid from the proceeds. Most of the contributing producers also supply some labour.

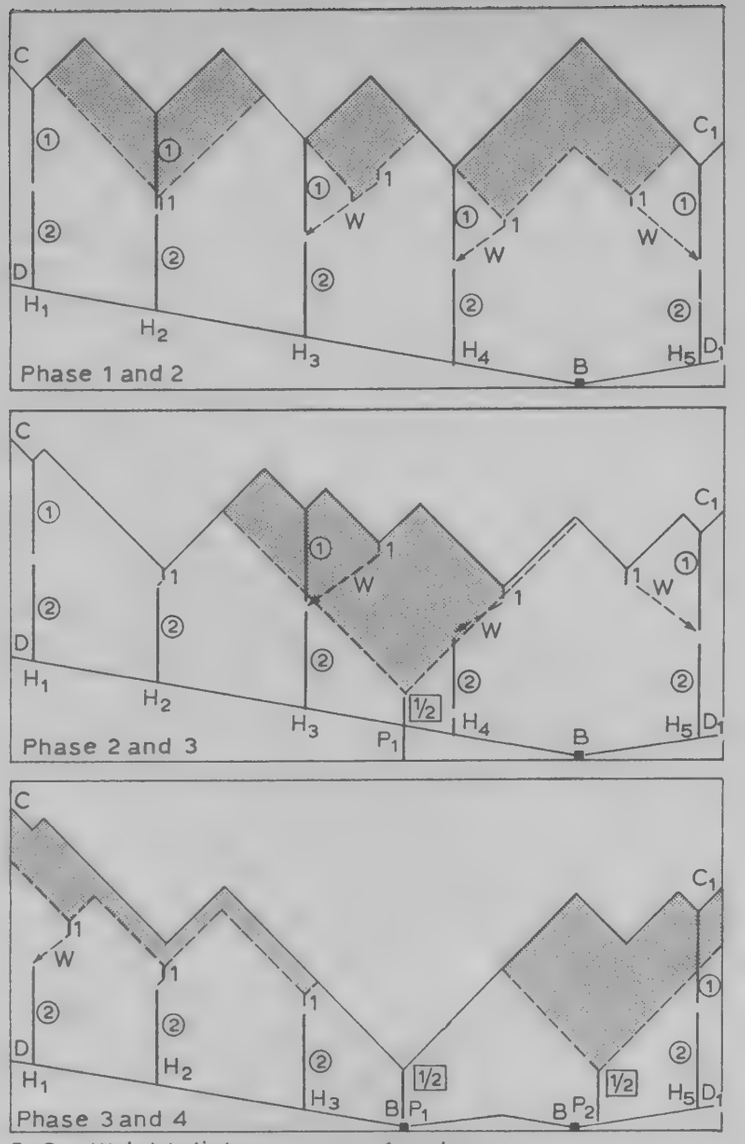


Fig. 4.10 A Weberian model of coffee processing; for a detailed note of assumptions see the original paper

I had the good fortune to be present in the field during the first three months' operation of this first central pulpery in 1965, a period which extended into the beginning of the annual flush season, which usually runs from March or May until about October. The whole three months' production was sold in one lot, totalling 1517 kg of parchment, at a price of 55 cents per kg, a premium of 11 cents per kg above the price then paid for individually processed coffee at the road-buying points. Not all this 55 cents per kg was received by the producer, because debts for cement and other materials used in the pulpery were first met from the undistributed returns, then the clerk and two operatives were compensated by individual contributions roughly graded according to the gross return of each producer. This operation was hard to follow, but receipts and contributions were fully recorded among 31 producers. Ignoring three very small producers whose contributions to the wage fund, though small, reduced their net return to under 30 cents per kg, the remaining 28 brought cherry equal to 763 kg of parchment, and obtained a mean net return of 49·8 cents per kg, with a range of 45·2 to 52·7 cents. Despite grading of contributions, these processing costs fell more heavily on the smaller producers. The 18 who got less than the mean net return produced a mean of only 19·8 kg of parchment each, while the 10 whose net return was equal to or above the average produced a mean of 40·8 kg each. There is little overlap in the two ranges.

A return to more dispersed processing: 1965–7

Iwagl central pulpery drew cherry from some members of each of two subclans in this initial period (Fig. 4.11). The co-operation needed, and the location of the central pulpery, arose from the long-standing association of the Councillors, progressive men who both live close to Iwagl. A number of men in both subclans stood out until after the first sale, when the large sums received and distributed impressed the doubters, and almost all Sunggwakani and many more Kamaniambugo joined the venture; the Tawiandumo pulping machine was removed to the central pulpery, though this was a move that did nothing to increase the capacity of the installation. Other Kamaniambugo, living at Mintima and on bad terms with the Councillor, refused to join and in 1966 persuaded the Administration officer serving the Council to propose allocation of Council funds to construct a second central pulpery for themselves; this remains grossly underused. A third central pulpery, also with Council support, was set up at Bamugl just west of the interclan boundary opposite Iwagl for the people of Numambugu clan. During 1966 the capacity of the Iwagl and Bamugl central pulperies was expanded by the construction of additional drying racks so that they could operate on a more continuous basis.

By mid-1967, however, many of these racks had collapsed without replacement, a large number of producers had returned to earlier forms of processing by hand-machine and subsequent men's house treatment, and even people with coffee close to the central pulperies were often using only the machines situated there, then taking wet beans back to the men's houses. The central pulperies were still used when large parcels of cherry were to be handled, especially during the flush, but the whole pattern of processing had again become more dispersed (Fig. 4.11).

This seeming retrogression followed an abrupt reduction in the coffee price late in 1966. Kundiawa Coffee Society was temporarily without funds, and private buyers who promptly moved in paid only 22 cents per kg in place of the former 44–55 cents. When the Society put its house in order and resumed buying, the price recovered only to 28–33 cents per kg. Furthermore, some enterprising Chimbu middlemen were at work, and the grower often realised no more than two-thirds or three-quarters of the factory price. At the same time men's house grouping received a new spurt as dispersed men were reconcentrated into new houses under the reported instruction of an over-zealous Administration officer. But it is possible to simulate such a dispersal, incorporating thoroughly rational explanations based on the work involved in carrying cherry – 5·5 times the weight of parchment – to the central pulperies, and on the comparative opportunity cost of the different modes of processing. On the basis of certain assumptions a simple Weberian model was constructed, and used to delimit 'equal cost' boundaries between processing points which, on a normative basis, should represent the cherry catchment areas of each (Fig. 4.10). These model boundaries are compared with the generalised actual movement in 1965 and 1967 in Fig. 4.11. Though there are real-world deviations, of which the most significant is the movement of much cherry to Iwagl from gardens

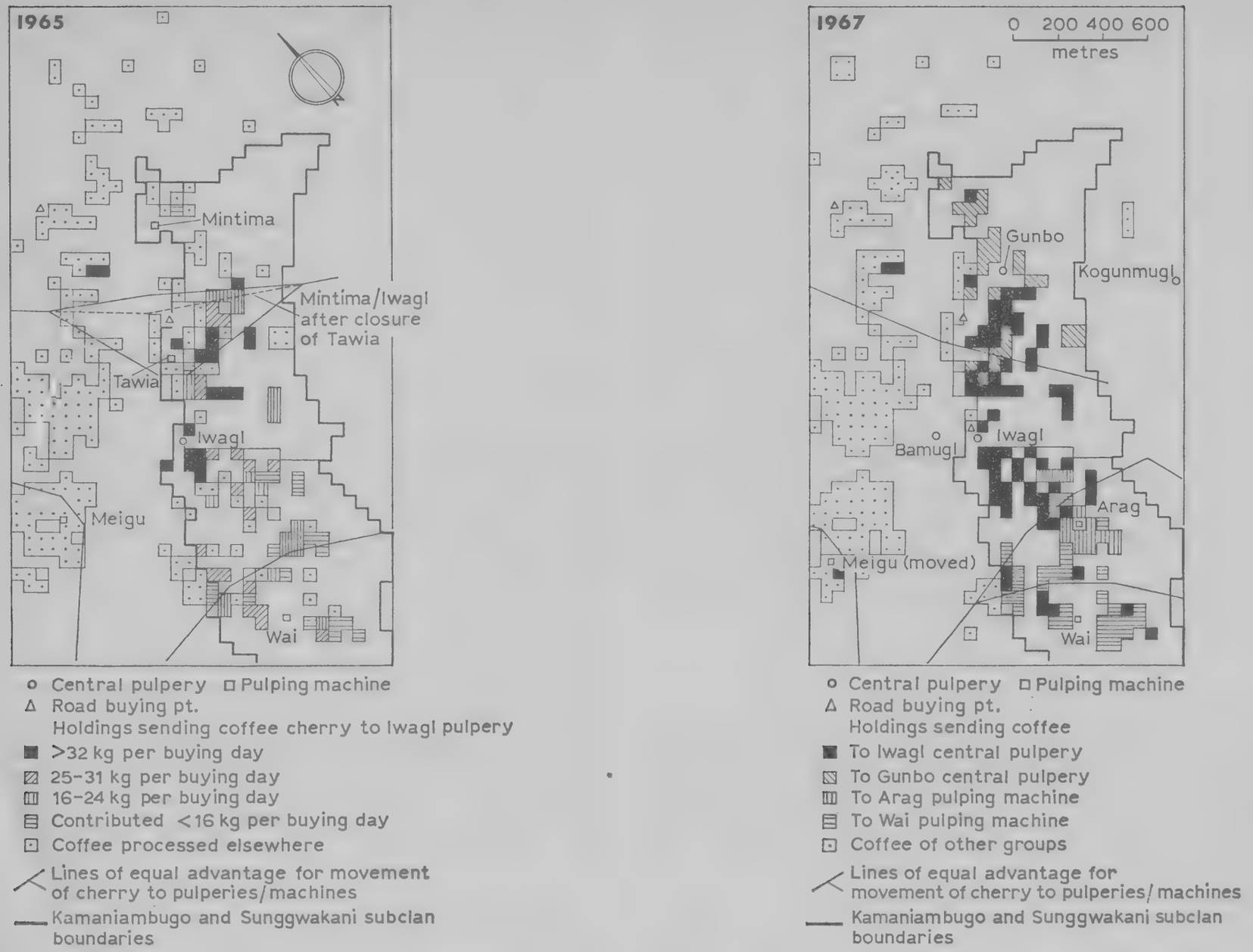


Fig. 4.11 Movement of coffee cherry from garden to processing point (from gridded data):

(*a*) Movement of cherry to Iwagl central pulpery, 20 Jan–16 Apr 1965
(*b*) Approximate movement of cherry to processing points in May 1967

Source: Based on information concerning individual producers, all of whose holdings, wherever located, are represented equally

close to Gunbo central pulpery, the general pattern shows some accordance with theoretical expectation. We may thus regard 1964–7 as a period during which, by experience, the Chimbu seem to have acquired a new perception of the spatial organisation of their coffee growing and processing.

Present uncertainties

The sharp fall in price in 1966 brought to an end the renewed optimism which followed the crisis of 1963. Coffee expansion was already slowing and was only small between 1965 and 1967. Knowledge that government policy was against further coffee expansion had filtered through, with many distortions, and generated gloom. The big new developments in tea cultivation in the Wahgi valley were being watched with some resentment. Coffee continued to provide a steady income, but it satisfied few of the progressives, and there was search for new innovations; a small-scale cattle project was attracting some interest, but was creating some severe problems at the ecological adaptation level. The growth of Kundiawa has improved demand for vegetables, and many women now take food to the twice-weekly market. A fair number of men, including many with good coffee holdings, had left the area to work – few now recruiting for plantation labour, but most going privately to the coastal towns to seek better-paid, though scarce, urban employment.

Despite the coffee retrogression there was little evidence of a return to traditionalism. Fundamental was the fact that major dietary changes had occurred since 1960, and have certainly gone much further between 1965 and 1967. Most families now have occasional meals of rice and tinned meat or fish; other store-bought foods, including bread and biscuits, have come increasingly into vogue. Many men like an occasional beer, and some drink more heavily. Manufactured cigarettes are increasingly preferred to twist-tobacco. The taste for a more varied diet, the general acceptance of washing, an improvement in the standard of store-bought clothing and its care and continuation of the inflation of traditional payments, especially for weddings, have created a dependence on the money economy that is now more than peripheral. The level of demand has continued to rise, though the means with which to satisfy it have, if anything, diminished since 1966. The one-time government ideal of a peasant population self-sufficient in basic foods, with a widely diffused but peripheral cash income, had perhaps been achieved, but this policy ceiling of an *ancien régime* in the erstwhile Department of Territories was far from providing any satisfaction level for Chimbu progressives who had followed the innovation path with energy for a decade.

Elsewhere in Chimbu, but not in this study area, the innovation path encountered another obstacle – that of shortage of available land for coffee. A number of Kamanegu have sought land near the Koro river from distant kin or have applied for settlement blocks farther west. Within the study area, Kamaniambugo subclan, with a density of 155 per sq. km, are the most short of land, and some men have sought land outside subclan territory for food crops; however, no one has yet complained of inadequate land on which to plant coffee.

EVOLUTION OF A CENTRAL PLACE

Decay of the old central place system

In terms of the spatial organisation of Chimbu society, the effects of successive innovations have gone far beyond coffee processing and men's house site-stabilisation. As the pace of coffee innovation has declined, modification of trading networks has accelerated, and has begun to be independent of its initial close link with the coffee innovation path. Again, I am concerned only with the local network. The decline of former long-distance trade routes north and south outward from the highlands, and their replacement by a new axial route along the centre of the highland valleys, is of concern here only in that the main highland road passes across the study area. We should, however, note that a modern central place for the whole Chimbu area, at Kundiawa 8 km distant, has been growing rapidly through the period, and especially since 1965.

There were three types of central place in the pre-contact man–environment system. The men's houses, where the men were nucleated, served as points of centralisation for the dispersed farming activities of the families of men living in the houses. Secondly, the spaces in front of the more important men's houses were places where members of a territorial group would foregather. In most years these two orders of central place were all the Chimbu had in current use. Periodically much more important, and notionally central to their catchment populations even when

not used, were the ceremonial ground sites. These were used two or three times a generation for major pig ceremonies, when the interpersonal network that cut across social groups, as well as the solidarity and competition of the social groups themselves, were recorded by massed gift-giving and group ceremony.

The ceremonial ground site at Mintima was among the largest in Chimbu, with a temporary population of about 1300 when last used in 1956. Like most of the larger grounds, Mintima stood on an interclan boundary. The centre was a pig shrine, or *bolum*, around which was a plaza where competitive dances, speech-making, mock hostility, pig killing, competitive display, butchering, cooking and giving of pigs took place. The whole ground was divided into segments of clan territory, subdivided between subclans. Each group had a men's house on its own territorial segment, fronting the central plaza. Women and pigs were located outward and behind them. In their layout and use, these grounds provided the Chimbu with something like a hardware iconic model of their society and its spatial organisation (Fig. 4.12).

The Mintima site remains in being, and is still used for some private or subclan minor functions; unlike some other ceremonial sites it has not been planted with coffee. But any further major ceremonies that take place seem likely to be of greatly modified type, and meanwhile modern central functions have aggregated elsewhere in the area.

Centrality, connectivity and the network

Centrality within the study area can be measured in a number of ways. The centre of garden distribution of a sample of 26 families, and the centre of all coffee, as measured by the central point of forces method, moved only within quite narrow limits between 1958 and 1965 (Fig. 4.12 (*c*)). Since male population distribution has always been more sensitive to centripetal forces than female population distribution, the network formed by all communal men's house sites ever occupied during the period, and the walking tracks joining them, was used to construct a so-called connectivity matrix, employing Pitts's 'short-path matrix', to obtain an 'index of net vertex connectivity' or centrality for each men's house site within the study area and its fringes, and secondly of each site occupied by a major men's house in 1958–9 over the whole territorial area of the population using Mintima ceremonial ground.[3] The latter exercise revealed the Mintima ground to be only on one edge of the area of 'most connected places' within its catchment: the former identified an axis of places of greatest centrality lying down the centre of the spur of land forming the core of the study area (Fig. 4.12 (*a*)).

Unrealistically assuming all tracks to be equal, a theoretical shortest path decomposition of the network into an ordered system focusing on the 'most connected' men's house site was prepared, and compared with a simplification of the actual network of very unequal tracks focusing on Iwagl central pulpery (Fig. 4.12 (*c*)). In this analysis only roads and tracks joining or passing close to men's house sites were included. A dense network of field paths, mainly of first and second order, joining gardens to women's houses and the latter to men's houses, was disregarded; unlike the higher-order tracks, this low-order network has been very unstable in location. Also disregarded are roads and footpaths leading out of the survey area, except where these form part of the local system of internal circulation. Both from the theoretical and actual results, it emerges that an axial highest-order path exists down the centre of the main spur, with apices at Mintima ceremonial ground, and at Iwagl–Bamugl. Central functions have aggregated at points on this axial path in a manner now to be described.

The emergence of Iwagl

As both demand and income levels rose when coffee began to come into bearing on a large scale in 1959, a number of native-owned trade stores was set up in the Chimbu area, and in the course of the period 1958–67 eight of these have been established in this study area, at points shown on Fig. 4.12 (*c*). Three outlying stores have failed. These trade stores carry a small stock of tinned meat, tinned fish, bread, tobacco, rice, sugar, salt, dripping, cigarettes, kerosene, utensils and clothing. By 1962 the turnover of the largest was already over $100 per month at cost prices, and for success in these enterprises it was necessary to locate close to roads, in order that goods could be delivered by Land-Rover.

After the establishment and initial success of the Iwagl central pulpery, there was a rapid increase in the central activities at Iwagl. Between 1965 and 1967 three further stores – one incorporating a bakery – were added to the existing trade store at this site and combined turnover approximated $100 per week. A medical

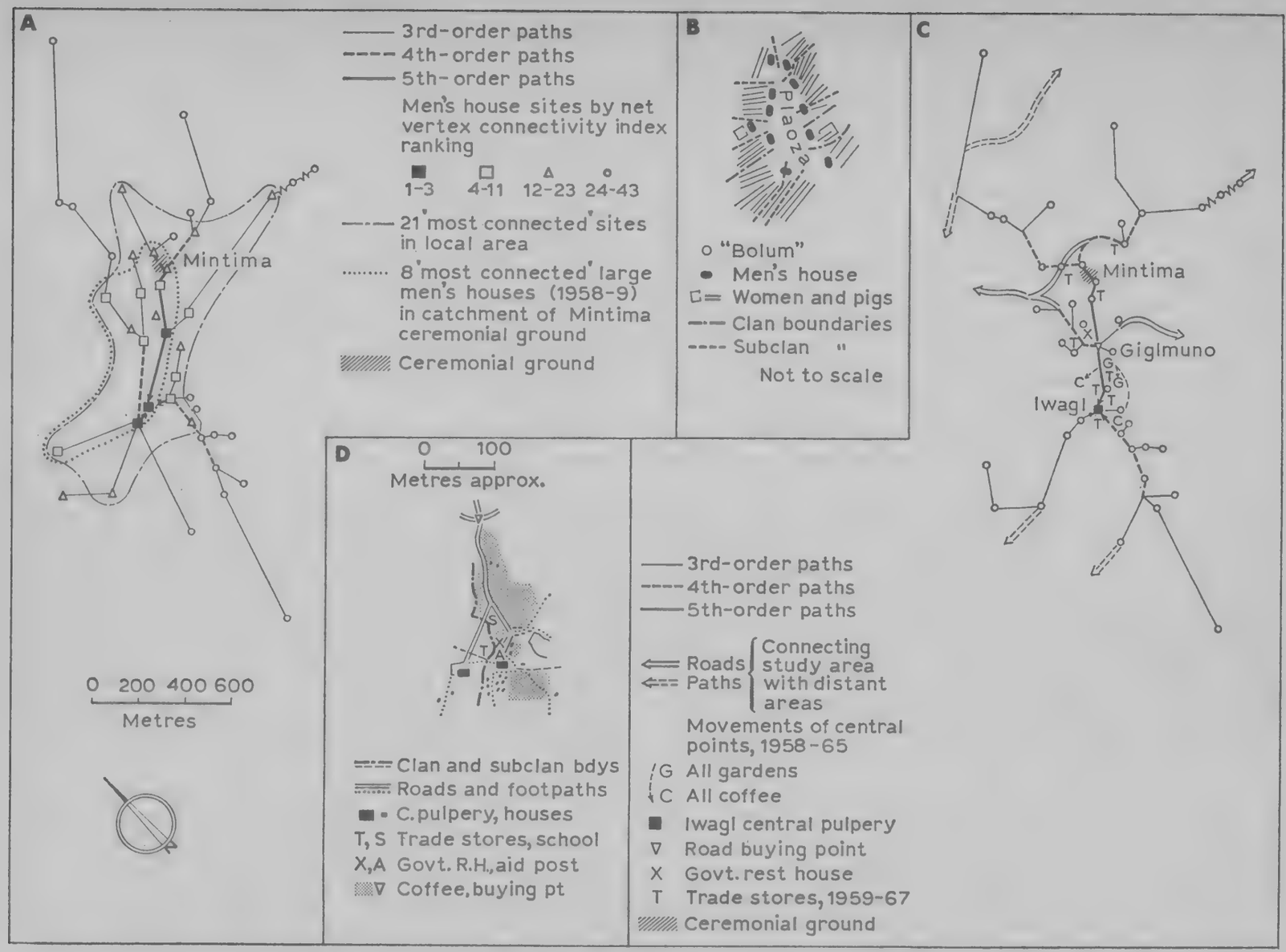


Fig. 4.12 Some elements in the system of internal circulation and centrality:

(*a*) Decomposition of the walking network by shortest path to the most connected men's house site, treating all tracks joining men's house sites as equal and ignoring first- and second-order paths from gardens and women's houses for simplicity; distribution of the most connected sites

(*b*) A diagram of Mintima ceremonial ground in 1956, from information obtained in 1959 and analogy with grounds elsewhere in use in 1959

(*c*) Decomposition of the walking network by shortest path to Iwagl, following the most used tracks from men's house sites; distribution of some consumer-oriented functions; movement of central points of distributions

(*d*) Iwagl in 1967 (from field survey and an air photograph)

aid post was established here, and in 1966 the Government Rest House was relocated at Iwagl from its former position higher up the axial path between the main road and the ceremonial ground. At the same time there was an aggregation of residential population into the Iwagl area. There were fewer than 20 adults living within 150 m of the site later occupied by the central pulpery in 1958; by 1967 there were between 50 and 60.

In all, there have been four central points along the axial path. The interclan boundary criss-crosses the path and all four central points have served members of both clans. The ceremonial ground at the upslope end was the first historically; later a Government Rest House site was established farther down the spur. The point where the axial path crosses the main highland road, at Giglmuno, became the main buying point, first for vegetables, later also for coffee; however, neither this point, nor a secondary buying point farther west on the road, have yet developed any other central functions. Fourth came Iwagl, initially as a trade store and the residence of the most influential of the local Councillors; then as a central pulpery serving only the area east of the interclan boundary. A second pulpery just west of the boundary was, as we have seen, badly located in relation to coffee. But other central functions rapidly aggregated at Iwagl; these served people from both sides of the interclan boundary, and have continued to expand even though the catchment of the central pulpery has dwindled. In so far as there is now a single central point for the population of the study area, it is Iwagl, located at the southern end of the axial path, but close to the central points of population, of all gardens and of coffee gardens alike. Its detailed location east of the interclan boundary and on the Kamaniambugo–Sunggwakani subclan border is determined by chance considerations. However, its general position is the one which, on any normative basis, we would deduce as the most likely for successful emergence of a central place internal to the system. Whether or not there will later be a movement upslope towards the point where the axial path crosses the main road, enabling the central place to serve more through traffic than at present, remains to be seen. The Naregu distrust the alleged thieving propensities of the through travellers, and consider the additional business to be gained as outweighed by the increased risk.

TOWARDS A SYNTHESIS

This discussion has by no means exhausted the consequences of the coffee and other innovations. Among consequences hardly touched on is increasing individualisation of economic activity, leading to growing differentials in wealth and the genesis of something like a plutocracy. Many women have acquired a personal cash income from coffee and vegetable sales, with important consequences in family structure, and in the position of women in public affairs. There is emerging a class of literate young adults; meanwhile very many young men are away at work or training. Rising affluence has also affected attitudes to the emerging national political system.

What I have tried to do is to trace out the consequences of the innovation of coffee, paying attention only to selected aspects, but at all times seeking to bear the whole system in mind. Okigbo succinctly traced the parallel consequences of an innovation in palm-oil milling in West Africa, and noted: 'An innovation introduces new factors into the social ecology. The society will adopt some of these factors and, in doing so, adapt itself both structurally and operationally in order either to resist the innovation or to accept its consequences. Social resistance may arise at the initial stage, giving rise to a re-evaluation of the economic change.'[4] In Chimbu, I have tentatively identified the events of 1963 as such a re-evaluation stage. The innovation was by then accepted and generally diffused, and its impact on the structure and operation of society and economy began seriously to be felt. This is perhaps the critical moment in an innovation path, and in Chimbu a rising level of demand – facilitated perhaps, by extensive outside experience gained in a decade of migrant labour – ensured that a majority passed this first critical point without halting or turning back. A fresh wave of innovations followed, of which some – outstandingly the formation of a new central place – occurred within the society. To achieve these changes, an ever-widening range of Chimbu perception of their total environment had to be reappraised; a revision of perception that began with crop ecology, proceeded to resource assessment and assessment of land needs and advanced to the spatial organisation of the farm, has now extended to a new evolving understanding of the spatial organisation of the economy as a whole. A sufficient number of

men has been carried through these difficult reappraisals by their continuously rising demand level, and their continuous expectation of being able to narrow the gap between present attainment and their satisfaction levels *within* the system. Further reappraisals, leading towards full individualisation and consolidation of land tenure, are likely to encounter greater resistance, for each new reappraisal digs deeper into the whole structure of society. To separate the individual economy from that of the group affects those aspects of the system concerned with group and interpersonal support, security and also the methods of capital formation which have served the innovation path up to now. Continued acceptance of new innovations, and successful adaptation to their consequences, demands not only that demand levels continue to rise, but also that expectations of achievement within the system do not fall too far behind. Despite the successes achieved, 'take-off to sustained growth' is by no means assured.

As development levels rise, more and more parallels emerge between this system and others in more complex societies. The emergence of Iwagl as central place, at the site of the first central pulpery, is analogous with the genesis of central places at the sites of grist mills in nineteenth-century Iowa. Innovation diffusion theory has obvious relevance at the spatial level, even though within the local system the trend seems to be from initial diffusion to later concentration, rather than vice versa. There are also cross-cultural parallels independent of the development process. I have elsewhere noted that pre-contact territorial distribution has analogues in north-west Europe, and that Chimbu settlement pattern is amenable to the same sort of analysis as that of Europe or Africa. Network analysis is as relevant to the analysis of the pre-colonial system as to the problem of central place emergence in the course of economic growth.

These analogues are in no sense curiosities. They arise from regularities in the processes tending to create equifinality in parts of initially very different systems. Models can be generated to simulate aspects of systems in 'primitive' societies as readily as for 'advanced' societies. Rigorous analysis may be harder to achieve among the former, but it is just as necessary. In our empirical task of interpreting the real world our basic methodological problem is the same in all societies: we cannot hope to establish causal relationships unless we make simplifying assumptions about other variables, that is unless we reduce the complexities of reality by generating some sort of partial model of the real world. If we first simplify our problem in this way, we may then search for regularities in terms of the causal interpretation of models such as the Weberian and central place models, or in terms of some unifying 'cause' such as a rising demand level. We cannot then pretend that we have dealt with all the complicating variables of the real situation, but we can at least draw causal inferences in terms of universal principles such as 'least effort' or 'satisficer' behaviour.

In all societies, this is our first task as empirical geographers: to reduce the level of locally unique noise, seek elements of regularity in terms of general principles of behaviour, investigate particular systems to see what they can contribute to general theory. Our local area, be it Chicago or Chimbu, is of generalist value first in terms of what its data can contribute to the general theory of the discipline. It is when we invert the analysis to apply general principles to the particular case that the locally unique complexities need to be taken fully into account, as we try to infer reasons for the deviations encountered. In this essay I have tried to take advantage of long and intimate familiarity with a small area and its people to illustrate these contentions: regularities emerge throughout the story, and can be analysed by simplifying local complexity. But to be able to do this confidently we must first immerse ourselves in the locally unique noise, since we are working in an unfamilar social milieu. This is the real meaning of the micro-study approach to human geography in the tropics that we have adopted of late years from Australia, and hopefully, this is the means by which we may be able to make our most useful general contribution to geography.

It is characteristic of human beings to live in societies. These societies are of many different kinds, whether small social groups like a street boys' gang or large nations like Great Britain, but they all have something in common: each society has its own social organisation and social patterns, its own way of life which is meaningful to those people who live in it.

This point is illustrated in the two otherwise very different societies described here. One is a street gang in an American slum, based on earlier Italian immigrants still not fully integrated into American society. The other is a band of wandering African hunters, people with few material or technological resources. Yet in each case these social groups form little societies of their own; in each we can find social patterns and social relationships and, as Whyte puts it, 'a highly organised and integrated social system'.

The social system of each of these societies has a number of different elements. Each is placed in a particular physical setting – the slum district of an American city, or the harsh terrain of the Kalahari Desert. In each there are economic problems that must be solved. There is a political aspect too: each group has its leader and means of making decisions and keeping conflict under control. There is also the psychological element; in each group there are personal as well as social needs to be met. In the two social systems here one can thus trace aspects which have been elaborated in previous readings: the psychological, political, economic and spatial.

A further element has been less apparent in earlier readings but is very marked in these two sociological extracts. This is the importance of social relationships. *Cornerville street gangs rest on a 'system of mutual obligations', while the Bushmen value co-operation and companionship and 'are sustained by a web of mutual obligation'.*

STREET CORNER SOCIETY

W. F. Whyte

INTRODUCTION: CORNERVILLE AND ITS PEOPLE

In the heart of 'Eastern City' there is a slum district known as Cornerville, which is inhabited almost exclusively by Italian immigrants and their children. To the rest of the city it is a mysterious, dangerous and depressing area. Cornerville is only a few minutes' walk from fashionable High Street, but the High Street inhabitant who takes that walk passes from the familiar to the unknown.

For years Cornerville has been known as a problem area, and, while we were at war with Italy, outsiders became increasingly concerned with that problem. They feared that the Italian slum dweller might be more devoted to fascism and Italy than to democracy and the United States. They have long felt that Cornerville was at odds with the rest of the community. They think of it as the home of racketeers and corrupt politicians, of poverty and crime, of subversive beliefs and activities.

Respectable people have access to a limited body of information upon Cornerville. They may learn that it is one of the most congested areas in the United States. It is one of the chief points of interest in any tour organised to show upper-class people the bad housing conditions in which lower-class people live. Through sightseeing or statistics one may discover that bathtubs are rare, that children overrun the narrow and neglected streets, that the juvenile delinquency rate is high, that crime is prevalent among adults and that a large proportion of the population was on home relief or Work Projects Administration during the depression.

In this view, Cornerville people appear as social-work clients, as defendants in criminal cases or as undifferentiated members of 'the masses'. There is one thing wrong with such a picture: no human beings are in it. Those who are concerned with Cornerville seek through a general survey to answer questions that require the most intimate knowledge of local life. The only way to gain such knowledge is to live in Cornerville and participate in the activities of its people. One who does that finds that the district reveals itself to him in an entirely different light. The buildings, streets and alleys that formerly represented dilapidation and physical congestion recede to form a familiar background for the actors upon the Cornerville scene.

One may enter Cornerville already equipped with newspaper

information upon some of its racketeers and politicians, but the newspaper presents a very specialised picture. If a racketeer commits murder, that is news. If he proceeds quietly with the daily routines of his business, that is not news. If the politician is indicted for accepting graft, that is news. If he goes about doing the usual personal favours for his constituents, that is not news. The newspaper concentrates upon the crisis – the spectacular event. In a crisis the 'big shot' becomes public property. He is removed from the society in which he functions and is judged by standards different from those of his own group. This may be the most effective way to prosecute the law-breaker. It is not a good way to understand him. For that purpose, the individual must be put back into his social setting and observed in his daily activities. In order to understand the spectacular event, it is necessary to see it in relation to the everyday pattern of life – for there is a pattern to Cornerville life. The middle-class person looks upon the slum district as a formidable mass of confusion, a social chaos. The insider finds in Cornerville a highly organised and integrated social system. . . .

DOC AND HIS BOYS

The members of the gang

The Nortons were Doc's gang. The group was brought together primarily by Doc, and it was built around Doc. When Doc was growing up, there was a kid's gang on Norton Street for every significant difference in age. There was a gang that averaged about three years older than Doc; there was Doc's gang, which included Nutsy, Danny and a number of others; there was a group about three years younger, which included Joe Dodge and Frank Bonelli; and there was a still younger group, to which Carl and Tommy belonged.

Since the Nortons, as I knew them, grew out of these earlier groupings, some historical background is necessary. The story of the evolution of the Nortons can best be told as Doc's story.

Doc was born on Norton Street in 1908. His mother and father, who came from the province of Abruzzi, were the first non-Genoese Italians to settle on the street. In a large family, Doc was the youngest child and his mother's favourite. His father died when he was a small boy. When he was three years old, infantile paralysis shrivelled his left arm so that it could never again be normal, but by constant exercise he managed to develop it until he was able to use it for all but heavy work. . . .

Doc was always sensitive about his arm, and he would not permit anyone to make allowances for his disability. He spent many hours at home shadow-boxing to develop speed and co-ordination.

Doc's most serious challenge came from Tony Fontana. As he told me:

> Tony was in my gang when we were kids together. He was a good fighter. When he entered the ring as an amateur, he started off winning three fights by knockouts. When he turned pro, he was still knocking them out. . . . At that time I was the leader of the gang. I was the tough guy. But he began to get fresh with me. One night he began pushing me around and talking big. I turned to him. I thought 'He must be tough. All those knockouts have got to mean something.' So after a while I said, 'I'm going up to bed.' I got undressed and went to bed, but I couldn't sleep. I put on my clothes and came down again. I said, 'Say that to me again!' He did and I let him have it – pow! . . . But he wouldn't fight me. Why? Prestige, I suppose. Later we had it out with gloves on the playground. He was too good for me, Bill. I stayed with him, but he was too tough. . . . Could he hit!

Doc told me these things only when I questioned him, and always, when he had finished telling about an incident in which he had 'flattened' some rival, he would half-apologise and say that he really wasn't so much, that he could hardly understand how such things had happened.

Every now and then there was a clash with some other gang, and a 'rally' resulted. . . .

> I don't remember that we ever really lost a rally. Don't get the idea that we never ran away. We ran sometimes. We ran like hell. They would come over to our street and charge us. We might scatter, up roofs, down cellars, anywhere. We'd get our ammunition there. Then they would go back to the other end of the street and give us a chance to get together again. We would come out one after another – they would never charge us until we were all out there and ready. Then we would charge

> them – we had a good charge. They might break up, and then we would go back to our end of the street and wait for them to get together again. . . . It always ended up by us chasing them back to their street. We didn't rally them there. We never went looking for trouble. We only rallied on our own street, but we always won there. . . .

Danny was his closest friend. As Doc told me:

> Danny lived on Stone Street near Norton. I remember now the day when he came over to our street, when he was a small kid. He was a greaser – spoke broken English. The fellows made fun of him, but I liked the kid from the start. I told him to come along with our gang and do the things we did. He stuck with us.

When the kids' gang broke up, Doc and Danny remained together, though they were not often seen on Norton Street.

Danny left school after the eighth grade to take a factory job. He supplemented his earnings there by organising a crap game in the washroom, and, in between jobs, he worked wherever there was labour trouble, for either side – 'whoever pays me'. Danny was powerfully built and well equipped by experience to fight in labour wars, but he did not relish this work. He fought for the money that was in it. In all my time in Cornerville, I never heard of Danny picking a fight with anyone.

With the passage of new labour legislation and government action against strike-breaking agencies, one of Danny's sources of income dried up. He had to fall back upon a crap game that he conducted in partnership with Mike Giovanni and Mike's brother, Terry.

Mike had been the leader of his kids' gang on King Street. He also had left school early for a factory job and was active wherever labour battles were being fought. Unlike Danny, he worked for only one side – the union. As he explained: 'Unionism is like a religion. You have those beliefs, and you have to stick to them.' In the last years of prohibition, factory work in his line became scarce, and Mike supported himself by running a crap game and a small speakeasy. He did not like the crap game, nor was it very profitable, for he refused the business of those who especially could not afford to lose. He thought he had 'the right connections' to provide protection for the speakeasy, but police raids forced him out of business. He opened a lunchroom, but that also proved unprofitable, since too much of the business was on a credit basis. While the lunchroom operated, it provided a social centre for Mike and his friends. Danny was a frequent visitor, and Doc spent some of his time there.

Long John, a young man from another part of Cornerville, took to hanging with Mike's crowd. He had been hanging with a particularly tough gang up to the time that his older brother was sentenced to life imprisonment for murder. Prodded by his mother, Long John began to worry about his own future. Danny and Mike advised him to break away from his former associates and hang with them. For some time they took care of his spending money and let him earn small sums by serving as watchman (looking out for the police) in their crap game. Then he got himself a factory job which provided sporadic employment throughout the year.

When Doc's kids' gang broke up, Nutsy was the only member who continued to spend all his spare time on Norton Street. Since he took up with the younger boys, Doc and Danny called him 'The King of the Kids'. Frank Bonelli became particularly attached to Nutsy. Joe Marco, known as Joe Dodge, was a good friend of both men. Carl and Tommy, who had belonged to a still younger group, now accepted Nutsy's leadership. Alec had gone to school with a younger brother of Joe Dodge, and he first took to hanging on Norton Street in order to be with Joe.

At this time Nutsy was a part-time postal employee, Frank was trying to get started in professional baseball and Joe had a highly paid but seasonal job in a quarry. Carl and Tommy both had steady factory jobs, and Alec had seasonal employment in the market district.

Besides Mike's crowd and Nutsy's boys, there were three other men who went to make up the Nortons as I knew them. Angelo Cucci, Fred Mackey (Macaluso) and Lou Danaro were all closely attached to Doc. Some years earlier, Fred's uncle had opened a grocery store on Norton Street and had placed Fred in charge part of the time. One day Danny got the boys to play a practical joke on him. They lined up in front of the counter and demanded protection money. Fred was panic-stricken until Doc took pity on him and explained the situation. Fred was so relieved that he

looked upon Doc as his benefactor and frequently sought his company, even after the store had been sold.

For several years Lou Danaro had worked for Mr Bacon, the headworker of the Norton Street Settlement, and had even lived in the house. The corner boys thought that he considered himself above them, and they would have nothing to do with him. Doc knew Lou's cousin very well. The cousins did not get along; Doc thought that was too bad, so, when he went out with Lou's cousin, he insisted that they get Lou to join them. In that way he struck up a friendship with Lou. When Lou finally broke with Mr Bacon and moved out of the settlement, his friendship with Doc made it possible for him to be accepted on the street corner.

Fred and Lou both lived in the suburbs, but they drove into Eastern City for their part-time jobs and into Cornerville to join Doc and his friends.

When Doc first met him, Angelo was exceedingly shy and had no friends. He spent most of his time at home practising the violin, which he hoped some day to play in a concert orchestra. When Doc accepted him as a friend, it was possible for Angelo to join the corner boys.

Close friendship ties already existed between certain of the men, but the Nortons, as an organisation, did not begin to function until the early spring of 1937. It was at that time that Doc returned to the corner. Nutsy, Frank, Joe, Alec, Carl and Tommy had a great respect for Doc and gathered around him. Angelo, Fred and Lou followed Doc in making the corner their headquarters. Danny and Mike were drawn to Norton Street by their friendship for Doc and by the location of their crap game, right next to 'the corner'. Long John followed Danny and Mike.

The men became accustomed to acting together. They were also tied to one another by mutual obligations. In their experiences together there were innumerable occasions when one man would feel called upon to help another, and the man who was aided would want to return the favour. Strong group loyalties were supported by these reciprocal activities.

There were distinctions in rank among the Nortons. Doc, Danny and Mike held the top positions. They were older than any others except Nutsy. They possessed a greater capacity for social movement. While the followers were restricted to the narrow sphere of one corner, Doc, Danny and Mike had friends in many other groups and were well known and respected throughout a large part of Cornerville. It was one of their functions to accompany the follower when he had to move outside his customary social sphere and needed such support. The leadership three were also respected for their intelligence and powers of self-expression. Doc in particular was noted for his skill in argument. On the infrequent occasions when he did become involved, he was usually able to outmanœuvre his opponent without humiliating him. I never saw the leadership three exert their authority through physical force, but their past fighting reputations tended to support their positions.

Doc was the leader of the gang. The Nortons had been Doc's gang when they had been boys, and although the membership had changed, they were still thought to be Doc's gang. The crap game and its social obligations prevented Danny and Mike from spending as much time with the Nortons as did Doc. They were not so intimate with the followers, and they expected him to lead.

Long John was in an anomalous position. Though he was five years younger than Doc, his friendship with the three top men gave him a superior standing. As Doc explained:

> It's because we've always catered to Long John. When we go somewhere, we ask Long John to go with us. We come up to him and slap him on the back. We give him so much attention that the rest of the fellows have to respect him.

Nevertheless, he had little authority over the followers. At this time he was accustomed to gamble away his week's earnings in the crap game, and this was thrown up against him.

There is an important social distinction between those who hold crap games and those who play in them. The game-holders enjoy something of the standing of businessmen; the 'shooters' are thought to be suckers. The Nortons as a group considered themselves above the crap-shooters' level, and at this time Long John was trying unsuccessfully to break away from the game.

In the spring of 1937 Nutsy was recognised informally as the superior of Frank, Joe and Alec, but his relations with a girl had already begun to damage his standing. A corner boy is not expected to be chaste, but it is beneath him to marry a girl who is 'no good'. Nutsy was going so steadily with this girl that marriage seemed a distinct possibility, and, reacting to the criticism of

his friends, he gradually withdrew from the gang. He did not again play a prominent role in the Nortons until towards the end of my stay in Cornerville, but in the spring and summer of 1937 he was still a man of moderate importance.

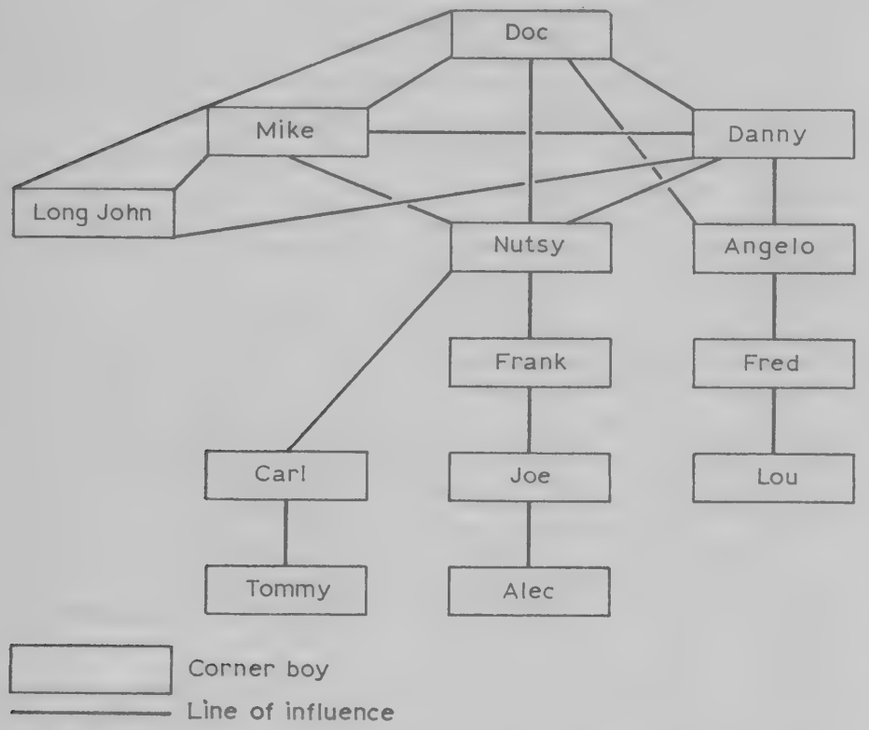


Fig. 5.1 The Nortons, spring and summer 1937. Positions of boxes indicate relative status

As the story gets under way, Doc was twenty-nine; Mike, twenty-nine; Danny, twenty-seven; Long John, twenty-four; Nutsy, twenty-nine; Frank, twenty-three; Joe, twenty-four; Alec, twenty-one; Angelo, twenty-five; Fred, twenty-five; Lou, twenty-four; Carl, twenty-one; and Tommy, twenty. Fig. 5.1 presents a picture of the relations between the men as they appeared at this time. For purposes of shorthand designation, I shall refer to the top four men as the 'leaders' and to the others as the 'followers'. If the special characteristics of Long John are borne in mind, this should not be confusing. . . .

The Nortons and the Aphrodite Club

In March 1938 the Nortons made the acquaintance of the Aphrodite Club girls. The club had a dozen members, most of them attractive, all of them well dressed, who met once a week in the Norton Street Settlement House with one of the social workers. The girls went to plays, held socials and each year saved up their dues for a trip to some point of interest.

In the winter of 1937–8 the Italian Community Club and the Aphrodite Club had become very friendly. They met in the settlement house on the same evening and sometimes gave their social activities together. The girls had a great respect for education and were anxious to make a good impression on the Community Club boys, although they thought some of them conceited. The men found the girls attractive, but some of the leading figures in the Community Club were anxious to make contacts with girls from the Italian Junior League and thus leave the Aphrodite girls and Cornerville behind them in their social contacts. While individual members were still mildly interested in some of the girls, by March the two clubs were slowly drifting apart.

Doc, Angelo, Lou and Fred knew the girls through being members of the Community Club, but up to this time the Nortons as a whole had no social contacts with the Aphrodite Club. To all outward appearances their attitude was hostile. In June 1937 I heard them discussing Carrie, one of the most attractive of the Aphrodite girls:

> *Nutsy:* She's a good-looking girl, but I don't like her.
>
> *Frank:* If you took three hours to make up your face, you'd be good-looking too.
>
> *Long John:* She has tough pins [legs]. Did you ever notice the pins on her? That's why she always wears such low dresses.
>
> *Joe Dodge:* She goes for anybody with a little money. She likes you if you have a car. I like to drive by her in my car and stick my nose in the air. . . . She's just an alley cat.

Except for the four men who belonged to the Community Club, the Nortons almost never set foot inside the settlement house. The girls moved in a different social orbit, and the Nortons considered them 'high-toned' and conceited. Still, they could not help finding them attractive.

As Doc told me:

> They had admired the girls for a long time, and they were always after me to fix it up with them. . . . Friday night the Community Club was going bowling after the meeting. They wanted me to go with them, but I stalled them off. I said I would be down in a little while. . . . Then I rounded up these boys [the Nortons] and told them we were going to bowl the girls. . . . I think the Aphrodite Club expected they were going to bowl with the Community Club. If I had asked them to bowl with this bunch, they would probably have refused because it's a tougher bunch. . . . But I just brought them down here and we bowled.

The evening was a great success. The two groups bowled together again on Saturday night, and on both occasions the bowling was followed by 'coffee-ands' at Jennings'.

Alec discussed this beginning with me: 'Before, we thought they were high-hat, and I guess they thought we were a bunch of rowdies. . . . Now I think they like our clique. We're cutting out the Community Club.'

Members of the two groups saw one another almost every night for a period of several weeks. This brought about important changes in the social life of the Nortons.

One night, only four days after the first bowling match, I was surprised to find Alec, Joe Dodge, Tommy and Long John playing cards at a table in the game room of the Norton Street Settlement House. Doc was at a table by himself, reading a magazine. I sat down with him to demand an explanation. He told me that the boys had wanted to go bowling but that he had not, so he told them to come into the settlement house and promised to get some of the girls to play cards with them. The girls were not in the house at the time, but, once inside, the boys played among themselves.

Except for Danny and Mike, the Nortons began coming into the settlement house almost every night to play cards by themselves or with the girls. Sometimes, when they were standing on the corner, the girls invited them to come in.

The social workers made no effort to keep the Nortons in the settlement. Miss Halloran, who was in charge of the game room, tried to treat them like the younger boys and girls in her charge. Corner boys are quick to notice the slightest sign of condescension, and Miss Halloran's attitude was painfully apparent. For several days the boys seemed to be obsessed with the task of denouncing Miss Halloran to one another in the strongest language they could command.

Long John, who had travelled the greatest social distance to go into the house, was the first one to drop out. After one particularly unpleasant encounter with Miss Halloran, he told the boys that he would never go in again. Two days later I was standing on the corner with him when it began to rain. Not knowing of his resolution, I suggested that we go into the settlement. He agreed, but, as he opened the door, we met Joe Dodge, who laughed at him and said, 'I thought you weren't going to come in here any more.'

Long John was embarrassed. We went out into the rain again, and he remarked thoughtfully, 'I think that everybody that goes in there thinks they're a little better than the next fellow.'

Two weeks after their first evening in the game room, all the Nortons had deserted the settlement.

From the beginning the boys took their activities with the girls very seriously. When they went bowling for the second time, Alec brought along a box of candy from his uncle's shop. The following Saturday he brought a large supply of pastry.

Doc told the boys that the Alluni sisters and their cousins had a summer camp on a lake some miles from the city. If the boys became friendly with the girls, they might be invited to spend the day in the country sometime during the summer. The possibility of taking the girls away from the Community Club was another inducement.

In a short time the Nortons did supersede the Community Club, but victory was achieved only by default. Tony Cardio was infatuated with Helen, the most attractive of the girls, but the other club members lost interest in the Aphrodite Club. However, since Tony was considered one of the two most conceited members, the Nortons could get almost a full measure of satisfaction from defeating him.

A week after the first meeting of the two groups, I asked Alec what he thought the association with the girls had done to the Nortons. He said: 'The boys get along better. There aren't so many squabbles any more.' At this time the Aphrodite girls and the Nortons met *en masse*. Alec commented:

> If I went out with them a few times, I could tell which one I liked. . . . But you start going with one girl, and you find you're going with a deadhead. It's tough. What are you gonna do? When they're all in a crowd, it's hard to shift around.

The boys had to proceed carefully. They could pay attention to Helen as much as they liked, because she was the prize of them all, but pairing-off with any of the others required a cautious preliminary survey of the situation.

A week later the first step in this direction was taken. Joe Dodge, Long John, Frank Bonelli and I were standing on the corner. Angelo Cucci encountered Alec farther down the street and told him that he had just seen three of the Aphrodite girls on their way to Jennings'. They had remarked that it would be a nice evening for a ride but that they did not believe that Joe Dodge owned the big car the boys all talked about. Alec walked up to us and took Joe aside. Joe then left us and walked across the playground to King Street. Frank, who had been watching attentively, turned to me and asked if I wanted to walk down to Jennings' with him. I said that I thought it was too early. Frank started off alone. Alec asked me if I wanted to walk to Jennings' with him. Long John said, 'You ain't bulling us. Why don't you tell the truth and say you're goin' for a ride?'

I asked Long John if he wanted to walk down with us. He refused but said that I should go.

As Alec and I walked down Main Street, he said that we were to meet Joe Dodge in his car, drive to Jennings', and take out the girls. 'But now Frank has started down to Jennings'. . . . He shouldn't do that.' There were two reasons for excluding Frank: he was shy with girls and he had no money to entertain them.

We went into Jennings' to talk to the three girls. One of them had to be home early, but she urged that the other two go. I persuaded Joe and Alec to leave me behind. By this time Frank had come in and sat down at a table by himself. As Joe and Alec went out with the girls, I joined Frank. A few minutes later Long John and Nutsy came in and joined us. Long John asked, 'What's the matter, Bill? Did they double-cross you?' I explained what had happened, but he said he did not like the way they had acted.

Frank and Alec had been the best of friends and had spent long hours together in the pastry shop. Now Alec became more friendly with Joe Dodge, and they began taking the girls out in Joe's father's car. Frank said to me:

> Let them go out with the girls. They've pulled a few fast ones. They say they're going to do a certain thing, and then you find them with the girls. They've done things that I would never do. . . . It's hard enough to make a friend. A girl you can meet any time. . . . It takes years to make a real friend.

The rift between Alec and Frank widened rapidly. Easter was the rush time at Alec's uncle's pastry shop, and Alec had promised to give Frank some of the extra work at this time. Frank said that Alec simply decided not to give it to him. Alec said that he went to get Frank, but Frank was very gruff and unpleasant so he refused to bother with him. In any case, Frank did not get this much-needed work, and he was bitter about it. Frank and Alec told their stories to Doc at different times. Alec complained that Frank was being ungrateful after he had done so much for him. Frank complained that Alec had been double-crossing him over the girls. Doc listened sympathetically but was unable to smooth things over. Joe Dodge's car and the Aphrodite girls had created too wide a gulf between them.

The activities of Alec and Joe made them unpopular with all the other Nortons – except Carl and Tommy, who had cars. They continued to hang on the corner, but for some time they were simply tolerated.

One Saturday night the Nortons were bowling with the girls. Ten men who had been members of the Sunset Dramatic Club were bowling on two adjoining alleys.

Danny and Mike came in late, sat by themselves, refused all invitations to bowl, and watched the proceedings with evident disgust. Danny told me: 'I don't like to bowl with the girls. There's no competition. . . . Then, when you get a touch hit, you can't say nothing. You got to watch what you say.'

Towards the end of the evening, Mike called Doc aside. He pointed to the Sunsets and said, 'We used to call them the Cream Puffs, but now, compared to you, they're the Lumberjacks.'

Doc laughed. Mike gave Danny a penny and Danny gave it to Doc, saying, 'Toss this up. See it you're a man. Heads, you're a man. Tails, you ain't.'

Doc took this good-naturedly. But then Danny called over Mario Testa of the Sunsets and told him to tell Doc that the Cream Puffs had now become the Lumberjacks. Mario laughed. Doc became angry. Danny said to him, 'I'll spot you 20 pins and I'll beat you. I spot all the girls 20 pins.'

Doc accepted the challenge. Danny bowled 104 against Doc's 84 and enjoyed himself thoroughly. Doc said he did not mind being beaten by 20 pins; the next time he might beat Danny by 20 pins. Nor did he mind being kidded about bowling with the girls. He said he became angry only when Danny brought the Sunsets into the argument. While he had once belonged to their club, he took pride in his position with the Nortons, and he was sensitive about anything that would make them appear in an unfavourable light compared with the Sunsets.

At Jennings', after the bowling, Doc left the girls and sat with Danny. Danny agreed that it had been a mistake to bring the Sunsets into the argument, and he apologised. Doc said that he did not want Mike and Danny to be left out of the bowling any more than they did.

On the following afternoon Danny and Mike stood on the corner telling Frank, Long John and me what they thought about our association with the Aphrodite girls. Danny wanted to know what we were getting out of it.

> If you want to go places with them, you got to have money, and none of you have got a dime, so forget about it.
>
> Alec says to me, 'I'll lay them all.'
>
> I tells him, 'You won't lay a one of them, and I'll bet money on it. If you lay one of them girls, you'll marry her. That's the only way you'll ever lay one of them.'

Long John said that he never had cared much for bowling with the girls in the first place. Frank said that he had enjoyed it at first but that now all the fun was gone out of it. Mike said that he and Danny would form a 'grievance committee' and would readmit some of us to membership in the Cornerville Bears (this being the name of a championship baseball team on which he had once played) if we would swear not to bowl with the girls any more. Frank made his promise. Long John jokingly said that he would never apply for membership, and Danny said that Long John would get back only over his dead body.

Mark Ciampa came along and took Frank for a ride in his car with Joe Dodge and Carl. Lou Danaro drove up, and Danny, Mike, Long John and I climbed into his car, picked up Doc at his house and drove out to Crighton, where we stopped at the bowling alleys. Danny and Doc chose sides. Doc chose Lou and Mike; Danny chose Long John and me. But then Mike objected. He wanted to be on Danny's side to defend the honour of the Cornerville Bears. I changed sides with Mike, and Doc, Lou and I took two out of three strings from Danny, Mike and Long John, largely due to the fine bowling of Doc, who finished well ahead of Danny and thus felt that he had gained his revenge for the humiliation of the previous evening. When the match was over, Doc asked if we were all readmitted to the Bears. Mike said that we were. In a spirit of good fellowship, we drove back to Cornerville.

Later, Doc discussed these developments with me:

> I enjoyed bowling with the girls at first. I hoped that Mike and Danny would fall in. When they didn't, I didn't enjoy it so much any more. . . . I knew they didn't like it. They said to me, 'It isn't right. The girls are taking all the alleys.' . . . You might say that there was a little clash between us about bowling with the girls, but you saw how it worked out. It wasn't really serious. We got together again right away.

Saturday night became men's night once more as the bowling season drew to a close. Social relations with the Aphrodite girls continued for some months but on a curtailed scale. The girls' summer camp was the main attraction which maintained interest after the peak of the group activities had passed. The boys drove out to the camp several times during July and August.

Alec was always boasting about his prowess with the women. Doc paid little attention to him, but the other boys felt that something should be done to put him in his place. One night in April they were kidding Alec, when, as Doc says, Alec challenged him:

> 'If you're such a great lover, I challenge you to show your stuff!'
>
> I said, 'Alec, I might not be as handsome as you are, and I don't have all the hair that you have, but I can outbull you any day.'

Alec says, 'No! No!'

'Well,' I said, 'I'm older now and I don't want to take a girl away from a man just to show I can do it.'

But then Danny says, 'Doc, I think you're slipping.'

Maron! When Danny says that, I must do something. He only said it to steam me up, but I said, 'All right, Danny, I pick Helen. Saturday night. You watch.' . . . Alec wasn't there to see it Saturday night. That was too bad. We were bowling one floor below the girls. I went up to see Helen, and I asked her to come down. I had something to tell her. In a few minutes she came down – by herself. She sat next to me all the evening, the only girl among all those fellows. Danny was impressed. Later he told me, 'Doc, you're still the great lover.'

Since Alec was not present, he remained unimpressed and continued to boast. A month later Danny was again urging Doc to put Alec in his place. First, Doc lectured him upon the objectionable character of his boasting. When this had no effect, he asked: 'Which one of those girls do you really fit with?' Alec said he fitted best with Mildred.

'All right, you take her out twice more, so you can fit with her real good, and then I'll take her away from you.'

Alec protested that it could not be done. Later Doc commented to me:

I didn't think I could do it, but I said it anyway. I was all steamed up. . . . After, Alec called me aside, and he told me he loved Mildred and wanted to marry her, so I should lay off. . . . I said, 'All right, Alec, I just wanted to hear you say that.' . . . I don't think he really loves her, but that's the screwy code around here. If he says he loves her, I have to leave her alone.

Since Alec was more active with the girls than anyone except Joe Dodge, it required the intervention of the leader to put him in his place. Several months later he proposed to Mildred, and, when she refused him, he lost interest in the Aphrodite girls. A year later he married another girl.

When Doc took up Alec's first challenge and began to 'bull' Helen, he realised that he was running some risks. It would have been easy to fall in love with her, and Doc had no money or job on which to get married.

When the girls saw Doc with Helen, the combination seemed natural. Dorothy, one of Helen's closest friends, often remarked that they were such an attractive couple. In April, Helen was sick. As Danny told the story:

Dorothy is always hinting about flowers. She says a couple of times to Doc, 'Helen is sick. Why don't you send her flowers?' . . . That steamed me up. She's stupid. Don't she know that Doc can't afford to send flowers? . . . Last night me and Long John decided we should send her some flowers in Doc's name. He tried to tell us not to do it, but this morning we went down to Vanderwater, the florist on Silverton Street. . . . He had orchids, three for $15 – that was too expensive. We got roses; we told him it was not for a sweetheart, it was for a sick friend. So he suggests tear roses. We paid $5 for a dozen tear roses. . . . We could get six dozen roses around here for that price, but if we sent the flowers from one of these florists they would have some greaser knock on the door to deliver them. Vanderwater has a nice truck, and they send a delivery boy around with a green uniform. . . . But what things we couldn't do with that $5.

Doc greeted this gesture with mixed emotions. He realised that it had helped his standing, but it would make Helen think that he was serious. Finally, he told Dorothy and Helen that the boys had done it on his behalf.

Shortly after that, when the fellows were in Jennings', one of the girls was kidding Doc about his reputation as the great lover and claiming that he was afraid to go out with her. As he told the story:

Those kids get my goat. They're innocent, and they want to act as if they knew it all. . . . All right, I said I would go out with her. But she said, 'First you must come to my party.'

I asked her, 'Who's going to be there?'

'Tony Cardio, Chick Morelli and Angelo Cucci,' she says.

'Who else?'

'Nobody else.'

That steamed me up. Danny, Long John and Frank were at the same table with me, and she didn't invite them. . . . I told her, 'No, I'm going some place that night.'

She says, 'That's not true. You just don't want to come.'

'All right,' I said, 'I don't want to come.'

And she steamed up. When she went back to her table, I turned to the boys. They were very depressed. I told them, 'Pay no attention to it, she's stupid. She's tactless.'

Though he was unable to protect his boys from such a social slight, Doc at least showed that his interests were with them.

Association with the Aphrodite girls, combined with the bowling activity, brought about important changes in the life of Long John. In the spring of 1937 he was gambling away all his money in crap games. In the autumn of 1937 he began to cut down on his gambling, and by winter he had given it up entirely. In the spring I said to him that it must have taken a lot of will-power to stay away from the crap game for so long. He shrugged his shoulders. 'You know what really kept me away this winter? Bowling!'

Long John's attitude towards women began to change as he drifted away from the crap game and began associating with the Aphrodite girls. Whenever he stopped to think of it, he resented their attitude of social superiority, but at other times he found it very pleasant to be with them. Although he never thought seriously of marrying one of them, he said to me: 'If I could just find some girl that I could really fall in love with, I would get married tomorrow. . . . I really mean that.'

In a short space of time Long John had moved from a tough corner to the more respectable corner of Norton Street, from the crap game to the bowling alleys, from the alleys into the company of a select group of girls, and, with them, even into the settlement house for a brief period. As Doc commented, 'it was a metamorphosis'.

The Nortons and the Aphrodite girls were brought together by Doc. When Danny and Mike wanted to break them apart, they concentrated particularly upon Doc. Two of the followers could have been left out without changing the group very much, but Danny and Mike held such important positions that the Nortons could not have continued to be the Nortons without them. Furthermore, they were Doc's closest friends, and, whenever he had to choose between them and the others, he chose them. Bowling with the girls had threatened to split the Nortons, and Danny and Mike acted upon Doc to re-establish the unity of the group. By autumn the two groups had drifted apart so that one could hear the Nortons expressing the same attitudes towards the girls that they had held before becoming acquainted. Only Alec, Joe Dodge and Fred Mackey chose, in effect, to remain with the girls, and their relationship to the Nortons became rather tenuous. Joe and Fred eventually married into the group.

Association with the girls was, like bowling, a means of gaining, maintaining or losing prestige in the group. As in bowling, Alec had to be kept in his place. It was essential to the smooth functioning of the group that the prestige gradations be informally recognised and maintained. . . .

CONCLUSION: THE GANG AND THE INDIVIDUAL

The corner-gang structure arises out of the habitual association of the members over a long period of time. The nuclei of most gangs can be traced back to early boyhood, when living close together provided the first opportunities for social contacts. School years modified the original pattern somewhat, but I know of no corner gangs which arose through classroom or school-playground association. The gangs grew up on the corner and remained there with remarkable persistence from early boyhood until the members reached their late twenties or early thirties. In the course of years some groups were broken up by the movement of families away from Cornerville, and the remaining members merged with gangs on nearby corners. But frequently movement out of the district does not take the corner boy away from his corner. On any evening on almost any corner one finds corner boys who have come in from other parts of the city or from suburbs to be with their old friends. The residence of the corner boy may also change within the district, but nearly always he retains his allegiance to his original corner.

Home plays a very small role in the group activities of the corner boy. Except when he eats, sleeps or is sick, he is rarely at home, and his friends always go to his corner first when they want to find him. Even the corner boy's name indicates the dominant importance of the gang in his activities. It is possible to associate with a group of men for months and never discover the family names of more than a few of them. Most are known by nicknames attached to them by the group. Furthermore, it is easy to over-

look the distinction between married and single men. The married man regularly sets aside one evening a week to take out his wife. There are other occasions when they go out together and entertain together, and some corner boys devote more attention to their wives than others, but, married or single, the corner boy can be found on his corner almost every night of the week.

His social activities away from the corner are organised with similar regularity. Many corner gangs set aside the same night each week for some special activity, such as bowling. With the Nortons this habit was so strong that it persisted for some of the members long after the original group had broken up.

Most groups have a regular evening meeting-place aside from the corner. Nearly every night at about the same time the gang gathers for 'coffee-and' in its favourite cafeteria or for beer in the corner tavern. When some other activity occupies the evening, the boys meet in the cafeteria or tavern before returning to the corner or going home. Positions at the tables are fixed by custom. Night after night each group gathers around the same tables. The right to these positions is recognised by other Cornerville groups. When strangers are found at the accustomed places, the necessity of finding other chairs is a matter of some annoyance, especially if no nearby location is available. However, most groups gather after nine in the evening when few are present except the regular customers who are familiar with the established procedure.

The life of the corner boy proceeds along regular and narrowly circumscribed channels. As Doc said to me:

> Fellows around here don't know what to do except within a radius of about three hundred yards. That's the truth, Bill. They come home from work, hang on the corner, go up to eat, back on the corner, up a show, and they come back to hang on the corner. If they're not on the corner, it's likely the boys there will know where you can find them. Most of them stick to one corner. It's only rarely that a fellow will change his corner.

The stable composition of the group and the lack of social assurance on the part of its members contribute towards producing a very high rate of social interaction within the group. The group structure is a product of these interactions.

Out of such interaction there arises a system of mutual obligations which is fundamental to group cohesion. If the men are to carry on their activities as a unit, there are many occasions when they must do favours for one another. The code of the corner boy requires him to help his friends when he can and to refrain from doing anything to harm them. When life in the group runs smoothly, the obligations binding members to one another are not explicitly recognised. Once Doc asked me to do something for him, and I said that he had done so much for me that I welcomed the chance to reciprocate. He objected: 'I don't want it that way. I want you to do this for me because you're my friend. That's all.'

It is only when the relationship breaks down that the underlying obligations are brought to light. While Alec and Frank were friends, I never heard either one of them discuss the services he was performing for the other, but when they had a falling-out over the group activities with the Aphrodite Club, each man complained to Doc that the other was not acting as he should in view of the services that had been done him. In other words, actions which were performed explicitly for the sake of friendship were revealed as being part of a system of mutual obligations.

Not all the corner boys live up to their obligations equally well, and this factor partly accounts for the differentiation in status among them. The man with a low status may violate his obligations without much change in his position. His fellows know that he has failed to discharge certain obligations in the past, and his position reflects his past performances. On the other hand, the leader is depended upon by all the members to meet his personal obligations. He cannot fail to do so without causing confusion and endangering his position.

The relationship of status to the system of mutual obligations is most clearly revealed when one observes the use of money. During the time that I knew a corner gang called the Millers, Sam Franco, the leader, was out of work except for an occasional odd job; yet, whenever he had a little money, he spent it on Joe and Chichi, his closest friends, who were next to him in the structure of the group. When Joe or Chichi had money, which was less frequent, they reciprocated. Sam frequently paid for two members who stood close to the bottom of his group and occasionally for others. The two men who held positions immediately below Joe and Chichi were considered very well off according to Cornerville standards. Sam said that he occasionally borrowed money from them, but never more than fifty cents at a time. Such loans

he repaid at the earliest possible moment. There were four other members with lower positions in the group, who nearly always had more money than Sam. He did not recall ever having borrowed from them. He said that the only time he had obtained a substantial sum from anyone around his corner was when he borrowed eleven dollars from a friend who was the *leader* of another corner gang.

The situation was the same among the Nortons. Doc did not hesitate to accept money from Danny, but he avoided taking any from the followers.

The leader spends more money on his followers than they on him. The farther down in the structure one looks, the fewer are the financial relations which tend to obligate the leader to a follower. This does not mean that the leader has more money than others or even that he necessarily spends more – though he must always be a free spender. It means that the financial relations must be explained in social terms. Unconsciously, and in some cases consciously, the leader refrains from putting himself under obligations to those with low status in the group.

The leader is the focal point for the organisation of his group. In his absence, the members of the gang are divided into a number of small groups. There is no common activity or general conversation. When the leader appears, the situation changes strikingly. The small units form into one large group. The conversation becomes general, and unified action frequently follows. The leader becomes the central point in the discussion. A follower starts to say something, pauses when he notices that the leader is not listening, and begins again when he has the leader's attention. When the leader leaves the group, unity gives way to the divisions that existed before his appearance.

The members do not feel that the gang is really gathered until the leader appears. They recognise an obligation to wait for him before beginning any group activity, and when he is present they expect him to make their decisions. One night when the Nortons had a bowling match, Long John had no money to put up as his side bet, and he agreed that Chick Morelli should bowl in his place. After the match Danny said to Doc, 'You should never have put Chick in there.'

Doc replied with some annoyance, 'Listen, Danny, you yourself suggested that Chick should bowl instead of Long John.'

Danny said, 'I know, but you shouldn't have let it go.'

The leader is the man who acts when the situation requires action. He is more resourceful than his followers. Past events have shown that his ideas were right. In this sense 'right' simply means satisfactory to the members. He is the most independent in judgement. While his followers are undecided as to a course of action or upon the character of a newcomer, the leader makes up his mind.

When he gives his word to one of his boys, he keeps it. The followers look to him for advice and encouragement, and he receives more of their confidences than any other man. Consequently, he knows more about what is going on in the group than anyone else. Whenever there is a quarrel among the boys, he hears of it almost as soon as it happens. Each party to the quarrel may appeal to him to work out a solution; and even when the men do not want to compose their differences, each one takes his side of the story to the leader at the first opportunity. A man's standing depends partly upon the leader's belief that he has been conducting himself properly.

The leader is respected for his fair-mindedness. Whereas there may be hard feelings among some of the followers, the leader cannot bear a grudge against any man in the group. He has close friends (men who stand next to him in position), and he is indifferent to some of the members; but, if he is to retain his reputation for impartiality, he cannot allow personal animus to override his judgement.

The leader need not be the best baseball player, bowler or fighter, but he must have some skill in whatever pursuits are of particular interest to the group. It is natural for him to promote activities in which he excels and to discourage those in which he is not skilful; and, in so far as he is thus able to influence the group, his competent performance is a natural consequence of his position. At the same time his performance supports his position.

The leader is better known and more respected outside his group than are any of his followers. His capacity for social movement is greater. One of the most important functions he performs is that of relating his group to other groups in the district. Whether the relationship is one of conflict, competition or co-operation, he is expected to represent the interests of his fellows. The politician and the racketeer must deal with the leader in order

to win the support of his followers. The leader's reputation outside the group tends to support his standing within the group, and his position in the group supports his reputation among outsiders.

The leader does not deal with his followers as an undifferentiated group. Doc explained:

> On any corner you would find not only a leader but probably a couple of lieutenants. They could be leaders themselves, but they let the man lead them. You would say, 'They let him lead because they like the way he does things.' Sure, but he leans upon them for his authority. Many times you find fellows on a corner that stay in the background until some situation comes up, and then they will take over and call the shots. Things like that can change fast sometimes.

The leader mobilises the group by dealing first with his lieutenants. It was customary for the Millers to go bowling every Saturday night. One Saturday Sam had no money, so he set out to persuade the boys to do something else. Later he explained to me how he had been able to change the established social routine of the group. He said:

> I had to show the boys that it would be in their own interests to come with me – that each one of them would benefit. But I knew I only had to convince two of the fellows. If they start to do something, the other boys will say to themselves, 'If Joe does it – or if Chichi does it – it must be a good thing for us too.' I told Joe and Chichi what the idea was, and I got them to come with me. I didn't pay no attention to the others. When Joe and Chichi came, all the other boys came along too.

Another example from the Millers indicates what happens when the leader and his lieutenant disagree upon group policy. This is Sam talking again:

> One time we had a raffle to raise money to build a camp on Lake Blank [on property lent them by a local businessman]. We had collected $54, and Joe and I were holding the money. That week I knew Joe was playing pool, and he lost three or four dollars gambling. When Saturday came, I says to the boys, 'Come on, we go out to Lake Blank. We're gonna build that camp on the hill.'
>
> Right away, Joe said, 'If yuz are gonna build the camp on the hill, I don't come. I want it on the other side.'
>
> All the time I knew he had lost the money, and he was only making up excuses so he wouldn't have to let anybody know. Now the hill was really the place to build that camp. On the other side, the ground was swampy. That would have been a stupid place. But I knew that if I tried to make them go through with it now, the group would split up into two cliques. Some would come with me, and some would go with Joe. So I let the whole thing drop for a while. After, I got Joe alone, and I says to him, 'Joe, I know you lost some of that money, but that's all right. You can pay up when you have it and nobody will say nothin'. But, Joe, you know we shouldn't have the camp on the other side of the hill because the land is not good there. We should build it on the hill.'
>
> So he said, 'All right,' and we got all the boys together, and we went out to build the camp.

Disagreements are not always worked out so amicably. I once asked Doc and Sam to tell me who was the leader of a corner gang that was familiar to both of them. Sam commented:

> Doc picked out Carmen. He picked out the wrong man. I told him why he was wrong – that Dominic was the leader. But that very same night, there was almost a fight between the two of them, Dominic and Carmen. And now the group is split up into two gangs.

Doc said:

> Sometimes you can't pick out one leader. The leadership may be in doubt. Maybe there are a couple of boys vying for the honours. But you can find that out.

The leadership is changed not through an uprising of the bottom men but by a shift in the relations between men at the top of the structure. When a gang breaks into two parts, the explanation is to be found in a conflict between the leader and one of his former lieutenants.

This discussion should not give the impression that the leader is the only man who proposes a course of action. Other men frequently have ideas, but their suggestions must go through the proper channels if they are to go into effect.

In one meeting of the Cornerville S. and A., Dodo, who held a bottom ranking, proposed that he be allowed to handle the sale of beer in the clubrooms in return for 75 per cent of the profits. Tony spoke in favour of Dodo's suggestion but proposed giving him a somewhat smaller percentage. Dodo agreed. Then Carlo proposed to have Dodo handle the beer in quite a different way. and Tony agreed. Tony made the motion, and it was carried unanimously. In this case Dodo's proposal was carried through, after substantial modifications, upon the actions of Tony and Carlo.

In another meeting Dodo said that he had two motions to make: that the club's funds be deposited in a bank and that no officer be allowed to serve two consecutive terms. Tony was not present at this time. Dom, the president, said that only one motion should be made at a time and that, furthermore, Dodo should not make any motions until there had been opportunity for discussion. Dodo agreed. Dom then commented that it would be foolish to deposit the funds when the club had so little to deposit. Carlo expressed his agreement. The meeting passed on to other things without action upon the first motion and without even a word of discussion on the second one. In the same meeting, Chris, who held a middle position, moved that a member must be in the club for a year before being allowed to hold office. Carlo said that it was a good idea, he seconded the motion, and it carried unanimously.

The actions of the leader can be characterised in terms of the origination of action in pair and set events. A pair event is one which takes place between two people. A set event is one in which one man originates action for two or more others. The leader frequently originates action for the group without waiting for the suggestions of his followers. A follower may originate action for the leader in a pair event, but he does not originate action for the leader and other followers at the same time – that is, he does not originate action in a set event which includes the leader. Of course, when the leader is not present, parts of the group are mobilised when men lower in the structure originate action in set events. It is through observation of such set events when the top men are not present that it is possible to determine the relative positions of the men who are neither leaders nor lieutenants.

Each member of the corner gang has his own position in the gang structure. Although the positions may remain unchanged over long periods of time, they should not be conceived in static terms. To have a position means that the individual has a customary way of interacting with other members of the group. When the pattern of interactions changes, the positions change. The positions of the members are interdependent, and one position cannot change without causing some adjustments in the other positions. Since the group is organised around the men with the top positions, some of the men with low standing may change positions or drop out without upsetting the balance of the group. For example, when Lou Danaro and Fred Mackey stopped participating in the activities of the Nortons, those activities continued to be organised in much the same manner as before, but when Doc and Danny dropped out, the Nortons disintegrated, and the patterns of interaction had to be reorganised along different lines.

SHARING, TALKING AND GIVING: RELIEF OF SOCIAL TENSIONS AMONG !KUNG BUSHMEN

Lorna Marshall

THIS paper describes customs, practised by the !Kung Bushmen in the Nyae Nyae region of South Africa, which help them to avoid situations that are likely to arouse ill will and hostility among individuals within the bands and between bands. Two customs which seem to be especially helpful and which I describe in detail are meat-sharing and gift-giving. I mention also the !Kung habits of talking, aspects of their good manners, their borrowing and lending, and their not stealing.[1]

The common human needs for co-operation and companionship are particularly apparent among the Nyae Nyae !Kung. Independent living, outside the band structure,[2] does not exist for individuals. Nuclear families do not live alone. The arduous hunting-gathering life would be impossible without the co-operation and companionship of a larger group. Moreover, in this society the ownership of the resources of food and water is organised through the headmen of bands, and individuals have rights to these

resources by being members of a band connected to the headman by some near or remote kin or affinal bond. The !Kung are also extremely dependent emotionally on the sense of belonging and on companionship. Their need to avoid loneliness, I think, is actually visible in the way families cluster together in their werfs and in the way people sit, often touching someone else, shoulder against shoulder, ankle across ankle. Comfort and security for them lie in belonging to their group, free from hostility or rejection.

Their security and comfort must be achieved side by side with self-interest and much jealous watchfulness. Altruism, kindness, sympathy or genuine generosity were not qualities which I observed often in their behaviour. However, these qualities were not entirely lacking, especially between parents and offspring, between siblings and between spouses. One mother carried her sick adult daughter on her back for three days in searing summer heat for us to give her medicine. !Naoka carried her lame son, Lame ≠Gao, for years. Gau clucked and fussed over his second wife, Khuan//a, when she was sick. When !U had a baby, her sister, Di!ai, gathered veldkos for her for five days. On the other hand, people do not generally help each other. They laugh when the lame man, Kham, falls down and do not help him up. !U's jealous eyes were like those of a viper when we gave more attention to her husband, ≠Toma, than to her because he was much more ill. And, in the extreme, there is the report from the 1958 expedition of an instance of callous indifference in one band and the probable abandonment of a dying old childless woman, an old aunt, when her sister, with whom she lived, had died.

Occasions when tempers have got out of control are remembered with awe. The deadly poisoned arrows are always at hand. Men have killed each other with them in quarrels – though rarely – and the !Kung fear fighting with a conscious and active fear. They speak about it often. Any expression of discord ('bad words') makes them uneasy. Their desire to avoid both hostility and rejection leads them to conform in high degree to the unspoken social law. If they do deviate they usually yield readily to expressed group opinion and reform their ways. I think that most !Kung cannot bear the sense of rejection which even mild disapproval makes them feel. They also conform strictly to certain specific useful customs which are like instruments for avoiding discord.

TALKING AND 'TALKS'

I mention talking as an aid to peaceful social relations because it is so very much a part of the daily experience of the !Kung and, I believe, usefully serves three particular functions. It keeps up good open communication among the members of the band; through its constantly flowing expression it is a salutary outlet for emotions; and it serves as the principal sanction in social discipline. Songs are also used for social discipline. The !Kung say that a song composed specifically about someone's behaviour and sung to express disapproval, perhaps from the deepest shadow of the werf at night, is a very effective means of bringing people who deviate back into the pattern of approved behaviour. Nevertheless, during our observations, songs were not used as much as talking. If people disapprove of an individual's behaviour, they may criticise him or her directly, usually putting a question, 'Why do you do that?', or they may gossip a bit or make oblique hints. In the more intense instances 'a talk' may ensue.

The !Kung are the most loquacious people I know. Conversation in a !Kung werf is a constant sound like the sound of a brook, and as low and lapping, except for shrieks of laughter. People cluster together in little groups during the day, talking, perhaps making artifacts at the same time. At night families talk late by their fires, or visit at other family fires with their children between their knees or in their arms if the wind is cold.

There always seems to be plenty to talk about. People tell about events with much detail and repetition and discuss the comings and goings of their relatives and friends and make plans. Their greatest preoccupation and the subject they talk about most often, I think, is food. The men's imaginations turn to hunting. They converse musingly, as though enjoying a sort of day-dream together, about past hunts, telling over and over where game was found and who killed it. They wonder where the game is at present and say what fat bucks they hope to kill. They also plan their next hunts with practicality. Women (who, incidentally, do not talk as much as men) gave me the impression of talking more about who gave or did not give them food and their anxieties about not having food. They spoke to me about women who were remembered for being especially quick and able gatherers, but did not have pleasurable satisfaction in remembering their hot, monotonous, arduous days

of digging and picking and trudging home with their heavy loads.

Another frequent subject of conversation is gift-giving. Men and women speak of the persons to whom they have given or propose to give gifts. They express satisfaction or dissatisfaction with what they have received. If someone has delayed unexpectedly long in making a return gift the people discuss this. One man was excused by his friends because his wife, they said, had got things into her hands and made him poor, so that he now had nothing suitable to give. Sometimes, on the other hand, people were blamed for being ungenerous ('far-hearted') or not very capable in managing their lives, and no one defended them for these defects or asked others to have patience with them.

As far as we know, only two general subjects are taboo in conversation. Men and women must not speak openly together of sexual matters except as they make jokes in the joking relationship. The !Kung avoid speaking the names of the gods aloud and do not converse about the gods for fear of attracting their attention and perhaps their displeasure. They told us their myths, but only when we asked them to. It is evidently not the habit of these Bushmen to recount the myths as they sit by their fires. They do not invent stories. They said they had no interest in hearing things that are not true and wonder why anybody has.

While a person speaks the listeners are in vibrant response, repeating the phrases and interposing a contrapuntal 'eh'. 'Yesterday', 'eh', 'at Deboragu', 'eh', 'I saw old /Gaishay.' 'You saw Old /Gaishay', 'eh, eh'. 'He said he had seen the great python under the bank.' 'EH!', 'The PYTHON!' 'He wants us', 'eh, eh, eh', 'to help him catch it.' The 'ehs' overlap and coincide with the phrase and the people so often all talk at once that one wonders how anyone knows what the speaker has said.

Bursts of laughter accompany the conversations. Sometimes the !Kung laugh mildly with what I would call a sense of humour about people and events, often they shriek and howl as though laughter were an outlet for tension. They laugh at mishaps that happen to other people, like the lions eating up someone else's meat, and shriek over particularly telling and insulting sexual sallies in the joking relationship. Individual singing of lyrical songs, accompanied by the *//guashi*, or snatches of the medicine music, the playing of rhythmical games, the ceremonial dances themselves occupy the evenings as well, but mostly the evening hours are spent in talk.

'A talk', as the !Kung call it, differs from a conversation or an arranged, purposeful discussion. It flares spontaneously, I believe from stress, when something is going on in which people are seriously concerned and in disagreement. I think that no formalities control it. Anyone who has something he wants to say joins in. People take sides and express opinions, accusing, denying or defending persons involved. I witnessed one 'talk' only, in 1952,[3] which . . . I should mention in this context. It occurred over a gift-giving episode at the time of !Nai's betrothal and involved persons in Bands 1 and 2 who were settled near together at the time. Khuan//a, the mother of /Gunda, !Nai's fiancé, had diverted a gift which people thought was making its way to Gao Medicine, the present husband of !Nai's mother. Instead of giving it to him at the time when an exchange of gifts was in order, she gave it to one of her relatives. !Nai's mother's sister, !U, sitting at her own scherm, began 'the talk' by letting it be known what she thought of Khuan//a, in a voice loud enough to be heard in the two werfs, a startling constrast to the usual low flow of talk. Di!ai, !Nai's mother, sitting with her shoulder pressed against her sister's, joined in. People in the werfs went to sit at each other's scherms, forming little groups who agreed and supported each other. From where they sat, but not all at once and not in an excited babble, they made their remarks clearly, with quite long pauses between. Some expressed themselves in agreement with !U as she recounted Khuan//a's faults and deviations, past and present. Khuan//a's family and friends, who had moved to sit near her, denied the accusations from time to time, but did not talk as much or as loudly as !U. Khuan//a muttered or was silent. !U said she disapproved of her sister's daughter marrying the son of such a woman, but would reconsider her position if Khuan//a gave the expected gift to Gao Medicine. The talk lasted about twenty minutes. At that point Khuan//a got up and walked away and the 'talk' subsided to !U's mutterings and others' low conversation. In a few days Khuan//a gave Gao Medicine a present, not the gift in question, but one which satisfied Gao, and, as they said, 'they all started again in peace'.

There is a third form of verbal expression which might be called a 'shout' rather than a 'talk', but as far as I know the !Kung have

no special name for it. It is a verbal explosion. Fate receives the heat of the remarks in a 'shout'.

We were present on two such occasions, one in 1952, the other in 1953. Both were in response to the burning of scherms. In both instances little children, whose mothers had taken their eyes off them though they were only a few feet away, had picked up burning sticks from the fire, had dropped them on the soft, dry bedding grass in the scherms and, at the first burst of flame, had sensibly run outside unscathed. On the first occasion, the two children, who were about three years old, were frightened and were soothed and comforted by their mothers and other relatives. They were not scolded. On the second occasion, Khuan//a, the two-year-old granddaughter of Old ≠Toma and /Gam, had set fire to her grandparents' scherm. She was not apparently frightened at all and was found placidly chewing her grandfather's well-toasted sandal. She was not scolded either. What was especially interesting was the behaviour of the Bushmen. On both occasions they rushed to the burning scherms, shouting all at once, in extremely loud, excited voices, volcanic eruptions of words. The men made most of the noise, but the women were also talking excitedly. No one tried to do anything, nor could they, for the scherms burned like the fiery furnace. I asked the interpreters to stand close to one person at a time and try to hear what he said. People were telling where they had been when the fire started, why they had not got there sooner. They shouted that mothers should not take their eyes off the children, that the children might have been burned. They lamented the objects which had been destroyed – all in the greatest din I have ever heard humans produce out of themselves. It went on for about eight or ten minutes in bursts, then tapered off for another ten. While Old ≠Toma's scherm was burning, he and his wife, /Gam, the great maker of beads, sat on one side weeping. After the shouting had subsided, a dozen or more people set about looking for Old ≠Toma's knife blade and arrow points and picking up what beads they could find for /Gam in the cooling ashes. The two instances of 'shouts' provided examples of the vehemence which vocal expression can have and vividly illustrated the !Kung way of venting emotion in words.

There is still another kind of talk, not conversation, which I consider to be an outlet for tension and anxiety. We happened to hear it only in relation to anxiety about food and do not know if other concerns sometimes find expression in this way. It occurs in varying degrees of intensity. It is a repeating of something over and over and over again. For instance, whether it is actually so or not, someone may be reiterating that he has no food or that no one has given him food. The remarks are made in the presence of other individuals, but the other individuals do not respond in the manner of a discussion or conversation. In an extreme instance we saw a woman visitor go into a kind of semi-trance and say over and over for perhaps half an hour or so in ≠Toma's presence that he had not given her as much meat as was her due. It was not said like an accusation. It was said as though he were not there. I had the eerie feeling that I was present in someone else's dream. ≠Toma did not argue or oppose her. He continued doing whatever he was doing and let her go on.

All these ways of talking, I believe, aid the !Kung in maintaining their peaceful social relations by keeping everyone in touch with what others are thinking and feeling, releasing tensions, and keeping pressures from building up until they burst out in aggressive acts.

ASPECTS OF GOOD MANNERS

In !Kung society good manners require that, when !Kung meet other !Kung who are strangers, all the men should lay down their weapons and approach each other unarmed. ≠Toma, leader of Band 1, approached us unarmed the first time we met. After we were accepted and given Bushman names, we were no longer strangers and we never observed the practice again.

Good manners require that visitors be received courteously and asked to sit by the fire. The woman whose fire it is may welcome the visitor by taking a pinch of the sweet-smelling *sa* powder, which she carries in a little tortoise shell hung from her neck, and sprinkling it on the visitor's head in a line from the top of the head to the forehead.

Good manners in eating express restraint. A person does not reveal eagerness or take more than a modest share. When a visitor comes to the fire of a family who are preparing food or eating, he should sit at a little distance, not to seem importunate, and wait to be asked to share. On several occasions we gave small gifts of corned beef to be shared with a group. The person who

received the food from us would take only a mouthful. Once an old headman only licked his fingers. The lump of food would be passed from one to another. Each would take a modest bite. The last person often got the most. I found it moving to see so much restraint about taking food among people who are all thin and often hungry, for whom food is a source of constant anxiety. We observed no unmannerly behaviour, no cheating and no encroachment about food. Although informants said that quarrels had occasionally occurred in the past between members of a band over the time to go to gather veldkos, and although we observed expressions of dissatisfaction, no quarrels of any kind arose over food during our observations.

Good manners in general should be inoffensive. Any behaviour which is likely to stir up trouble is regarded with apprehension and disapproval by the !Kung. In view of this, the joking relationship has its interesting side. Men and women who have the joking relationship insult each other in a facetious way and also point out actual faults or remark on actual episodes which embarrass a person. Everyone joins in the uproarious, derisive laughter. All this is joking and one should not take offence. The !Kung say this teaches young persons to keep their tempers.

In contrast to the joking is their care in other aspects of conduct to avoid giving offence. ≠Toma said, for instance, that if he were forming a hunting party and a man whom he did not want asked to join him, he would be careful to refuse indirectly by making some excuse and try not to offend the man.

Gossip which can stir up trouble is discouraged. People do gossip but usually discreetly, in low voices, with near and trusted relatives and friends. It is best to mind one's own business, they say.

People are expected to control their tempers and do so to a remarkable degree. If they are angry, aggrieved or frustrated, they tend to mope rather than to become aggressive, expressing their feelings in low mutters to their close relatives and friends. ≠Toma told us that he had lost his temper twice when he was a young man and on one occasion had knocked his father down. On the other he had pushed his wife into hot ashes. It had so frightened him to realise that he could lose control of himself and behave in this violent way that, he said, he had not lost his temper since.

MEAT-SHARING

The !Kung custom of sharing meat helps to keep stress and hostility over food at a low intensity. The practical value of using up the meat when it is fresh is obvious to all and the !Kung are fully aware of the enormous social value of the custom. The fear of hunger is mitigated; the person one shares with will share in turn when he gets meat and people are sustained by a web of mutual obligation. If there is hunger, it is commonly shared. There are no distinct haves and have-nots. One is not alone.

To have a concept of the potential stress and jealousy which meat-sharing mitigates in !Kung society one has only to imagine one family eating meat and others not when they are settled only five or ten feet apart in a firelit werf and there are no walls for privacy. The veld does not hide secret killing and eating because actions are printed in its sands for all to read. The idea of eating alone and not sharing is shocking to the !Kung. It makes them shriek with an uneasy laughter. Lions could do that, they say, not men.

The custom of meat-sharing applies only to the big game animals. Small animals the size of duikers or smaller, birds, tortoises, lizards, grasshoppers, snakes and ostrich eggs are picked up incidentally and belong to the person who shoots, snares, finds or catches them. He may keep these for himself and his immediate family and may share them as he chooses, as if they were veldkos. ≠Toma says that, if there is only a little, he and his family eat a meal and give a little to anyone who happens to be near by at the time.

Meat-sharing applies to the big animals which are deliberately hunted by organised hunting parties. In the Nyae Nyae region they are kudu, gemsbok, wildebeest, hartebeest, springbok, warthogs and ostrich. Eland are found less commonly and buffalo rarely. The Nyae Nyae hunters managed to shoot giraffe occasionally. Giraffe are royal game by South-West Africa law, but the Nyae Nyae !Kung had not known this. When they were told, they called them 'Morrisi cows'. (Mr Robert Morris was the Commissioner of Native Affairs at Runtu.)

The system by which meat is shared begins with the hunters, but the composition of the hunting party is not a matter of convention or of anxious concern. Hunting parties are freely organised by men who like to hunt with each other. No categories of kin or

affines are prohibited from hunting together, whether they have a joking relationship or not or practise the sitting and speaking taboo. Men from different bands may hunt together.

A father has authority over his sons and sons-in-law and could tell them to go hunting or to accompany him and would expect them to obey. In other relationships participation in a hunting party is voluntary. Any man may instigate a hunt and may ask others to join. No one is formally in command of a party, unless he is the father with his sons or sons-in-law, but often an informal kind of leadership develops out of skill and judgement and the men fall in with the plans and suggestions of the best hunter or reach agreement among themselves somehow.

Hunting parties are usually composed of from two to four or five men. One hunter alone is at a disadvantage. A man cannot leave the meat unguarded from lions, hyaenas, jackals and vultures while he goes to bring help from the band to carry it, nor can he carry it alone. Only occasionally are parties larger than four or five. For two parties to hunt in different directions would be more advantageous than to form a larger single party.

When the kill is made the hunters have the prerogative of eating the liver on the spot and may eat more of the meat until their hunger is satisfied. If they are far from the band, they may eat the parts that are especially perishable or most awkward to carry, like the head, and they sometimes eat the cherished marrow. They then carry the animal to the band, either in its parts, bones and all, or, if the animal is very big, leaving most of the bones and cutting the meat into strips which dry in the sun to biltong and are more easily carried on the carrying-sticks. The blood is carried in bags made of the stomach, rumen and bladder.

The gall bladder and testicles are discarded at the kill. Eventually the hooves are thrown away and usually the picked bones and horns. (The !Kung make only a few artifacts of bone and horn which last a long time and seldom need replacing.) Sinews are kept for making cord. The hide would be skinned off whole and tanned if it were suitable for a kaross and someone wanted a new kaross at the time. Otherwise the hide is dried, pounded up and eaten. Hides are quite tasty. Feet are picked of every tissue; gristle is dried and pounded. Soft parts, such as the foetus, udder, heart, lungs, brains and blood, are often given to old people with poor teeth. Intestines are enjoyed and desired by all. The meat of the rump, back, chest and neck is highly appreciated. Nothing is wasted; all is distributed.

The owner of the animal is the owner of the first arrow to be effectively shot into the animal so that it penetrates enough for its poison to work. That person is responsible for the distribution. The headman of the band is not the one to distribute the meat by virtue of his headmanship, unless he happens to be at the same time the owner of the arrow. The owner may or may not be one of the hunters.

Hunters have arrows which they acquire in three different ways. Each man makes arrows for himself, shaping the points (usually now of metal, but still possibly of bone or wood) with some slight distinction so that he will know them from the arrows of other men. Secondly, arrows are given as gifts. The man who made the arrow or who himself acquired it as a gift, may give it to someone else, either a man or a woman, consanguineous kin, affine or friend. Thirdly, people lend arrows to one another. The status of the arrow plays its part in the distribution of the animal killed with it. There is much giving and lending of arrows. The society seems to want to extinguish in every way possible the concept of the meat belonging to the hunter.

A hunter chooses which arrow he will use. The owner of the arrow – who *ipso facto* owns the animal – may therefore be the hunter himself, who has chosen to use an arrow he made or one that was given him, or he may be a person who lent the arrow to the hunter.

There may be several hunters in the hunting party and several arrows in the animal, but this seems to cause no confusion or conflict. Every arrow is known, of course. The hunters may see which first penetrates effectively so that its poison could account for the kill. But I think that often it is arranged beforehand who will own the animal. A man asking another to accompany him might say, 'Come and help me get a buck.' Or 'Old Gau lent me an arrow and asked me to hunt for him. You come too.'

I think there is little or no dissension as to who owned the animal because it is not a cause for great stress; each hunter gets a share of the meat anyway. I think also that a man wants sometimes to be the owner of the meat in order to start the distribution off in the direction of his own kin and affines, but that one is also content sometimes not to have the onus of the main distribution.

The first distribution of the animal is made in large portions usually to five or six persons. They are the owner of the arrow, the giver of the arrow (if the arrow was not one the owner had made himself) and the hunters. The meat, always uncooked in the first distribution, is given on the bone, unless the animal is large and the meat has been cut into strips at the kill.

In a second distribution the several persons who got meat in the first distribution cut up their shares and distribute them further. This meat also is given uncooked. The amounts depend on the number of persons involved, but should be as much as the giver can manage. In the second distribution close kinship is the factor which sets the pattern of the giving. Certain obligations are compulsory. A man's first obligation at this point, we were told, is to give to his wife's parents. He must give to them the best he has in as generous portions as he can, while still fulfilling other primary obligations, which are to his own parents, his spouse and offspring. He keeps a portion for himself at this time and from it would give to his siblings, to his wife's siblings, if they are present, and to other kin, affines and friends who are there, possibly only in small quantities by then.

Everyone who receives meat gives again, in another wave of sharing, to his or her parents, parents-in-law, spouses, offspring, siblings and others. The meat may be cooked and the quantities small.

Visitors, even though they are not close kin or affines, are given meat by the people whom they are visiting. This is a strongly felt social rule. Visitors may receive small quantities of cooked meat, which it like being asked to dinner.

Name relatives often receive generous portions of meat because they have the same name as the giver or because their names associate them with his close kin, but this seems to be more a favour than an absolute rule. ≠Toma said there were far too many men named ≠Toma for him to give them special consideration.

It ends in everybody getting some meat.

In the later waves of sharing, when the primary distribution and primary kinship obligations have been fulfilled, the giving of meat from one's own portion has the quality of gift-giving. !Kung society requires at this point only that a person should give with reasonable generosity in proportion to what he has received and not keep more than an equitable amount for himself in the end, and that the person who receives a gift of meat must give a reciprocal gift some time in the future. Such factors as the category of relationship of kin, affines or non-related persons, the joking relationship, status and band affiliation impose no pattern on this giving. People give to whom they please, this time to some, next time to others, more generously or less generously according to their own (to us often unfathomable) reasons. We are certain, however, that the motives are the same as in gift-giving in general: to measure up to what is expected of them, to make friendly gestures, to win favour, to repay past favours and obligations, and to enmesh others in future obligation. I am sure that when feelings of genuine generosity and real friendliness exist they would also be expressed by giving.

In the parts and quantities of meat first distributed or given in the later waves of sharing there seemed to be no taboo and no rigidity or rule.

The distribution of an eland which was killed by Gao Beard, headman of Band 2, will serve as an example. There were more than a hundred !Kung at Gautscha at the time of the hunt. Both Gautscha bands (Bands 1 and 2) were present. Bands 3, 4 and 7 were visiting and there was a sprinkling of visitors from other bands.

The hunting party was composed of four men: Gao Beard; his first wife's brother, //Ao, and his own brother, /Qui, both of whom lived with him in Band 2; his brother-in-law, !Naishi, the husband of his sister, //Kushay. They were visiting at the time.

The party had hunted for eight days without success in heat so exhausting that they had to lie covered with sand through the middle part of the day.

!Qui was the first to see the eland. It was a huge one. As /Qui had been asked to come on the hunt by his brother to help him, he told Gao Beard where the eland was and did not shoot at it himself.

Two boys joined the men to track the eland after it was shot. The party tracked it for three days and then found it dead from the poison. They cut up the meat and brought it to the werf at Gautscha, which was two days' travel away. The hunt had lasted thirteen days in all.

Gao Beard was himself the owner of the arrow. The arrow had been given by one person to another five times. /Gao Music (Band 1), who made it, gave it to his sister, /Goishay, who gave it to her

husband, ≠Gao, headman of Band 3. He gave it to his brother, Gao, also of Band 3, who gave it to his wife, Di!ai. Di!ai gave it to Gao Beard, her brother, who shot the eland with it and who was responsible for the distribution of the meat.

Gao Beard first gave meat to the hunters who helped him, as was the custom. To !Naishi he gave a forequarter and to //Ao a forequarter and the head. (The hunters usually received a forequarter or hindquarter or an equivalent amount. The head was an extra gift.) The two boys who helped track got nothing – because their fathers would give them some, we were told. To our astonishment, /Qui was given nothing. Gao Beard explained that his brother would eat from his pot. (Actually he might eat more in this way as he would not have to share that cooked meat with anyone but his wife and child.)

Gao Beard's sister, Di/ai, who had given him the arrow, received the meat of the back and throat and the intestines.

Gao Beard kept the meat of the neck for himself and in a second wave of distribution gave the rest of the meat as follows:

To his first wife, //Kushay, he gave both hindquarters, the meat of the chest, the lungs, part of the liver and one hind foot. To his mother he gave the meat of the belly and one hind foot. To his sister, //Kushay, and to /Gasa, the wife of his brother, /Qui, he gave one front foot each.

The amount given to his first wife was enormous. In addition to giving to her co-wife, also named //Kushay, and their children, she was to give a large portion of it to her parents, who lived with them in Band 2. (On other occasions the man had given directly to his wife's parents – not through his wife.) When the meat was cut up, //Kushay gave to her father, Old ≠Toma, four bundles of strips of clear raw meat (it was somewhat dried by then). There were about ten or twelve strips to a bundle, about 2½ to 3 ft long. We guessed the weight in all to be about 60 to 70 lb. She gave meat also to her two younger brothers, /Gao and /Gunda. (Her other brother, //Ao, who had been one of the hunters, had got his share from Gao Beard.) She gave to her co-wife's father (≠Gao, Band 4) and to Old /Gaishay, her co-wife's MoFa, headman of Band 4. She then gave to six other persons, all in Band 1. They were her cousin, Old Gau (her FaSiSo) and his two daughters, and two other cousins (FaSiDas), Di!ai and !U, and !U's husband, ≠Toma.

The giving of raw meat went on. Old ≠Toma gave to eighteen people: his wife, six affines, three kin, two name relatives (i.e. a visitor because his father's name was ≠Toma, and Old /Gasa because her deceased husband's name was ≠Toma) and six other persons. Telling us the reasons for giving to the last six, whose kin or affinal connections with ≠Toma, if any, were so remote we did not bother to trace them, he said of one, 'He is an old man whom I like in my heart', and of another, 'He was hungry for meat.' In the end Old ≠Toma gave some of the meat back to his daughter, //Kushay.

Di!ai (Band 3), the giver of the arrow, gave raw meat to six persons: her husband's mother, brother and two sisters (in Band 2), a visitor who was her HuSiHuBr, and her mother's brother, headman of the visiting Band 7.

Persons who had by this time received substantial amounts of raw meat began giving to others. We recorded 63 gifts of raw meat. Doubtless there were more. After the raw meat was given, there was further sharing of the individual's portions, cooked or raw, with parents, offspring, spouses and other persons who live together.

Meat is not habitually cooked and eaten as a family meal among the !Kung. When an individual receives a portion of meat he owns it outright for himself. He may give and share it further as he wishes, but it never becomes family or group property. The men, women and children may cook their pieces when and as they wish, often roasting bits in the coals and hot ashes and eating them alone at odd times. Or someone may start a big pot boiling and several people will bring their pieces to put into it at the same time, each taking his own piece out when it is cooked.

The sense of possessing one's own piece personally is, I believe, very important to the !Kung. It gives one the responsibility of choosing when to eat one's meat and struggling with hunger as best one can when it is finished, without occasion or excuse for blaming others for eating more than their share.

It has often been reported that, when Bushmen kill an animal, they gorge themselves until they can hardly walk. We have seen the Nyae Nyae !Kung eat hearty meals of meat when they have been long without, but nothing more than we considered a normal amount. They hang meat in the bushes to dry and can keep it for some time. It is not uncommon for them to eat quite sparingly

and save bits for a coming journey or against a future day of hunger.

The !Kung are quite conscious of the value of meat-sharing and they talk about it, especially about the benefit of the mutual obligation it entails. The idea of sharing is deeply implanted and very successfully imposes its restraints. To keep meat without sharing is one of the things that just is not done.

GIFT-GIVING

The custom of gift-giving comes, in my opinion, second only to meat-sharing in aiding the !Kung to avoid jealousy and ill will and to develop friendly relations. !Kung society puts considerable emphasis on gift-giving. Everything a person has may have been given to him and may be passed on to others in time. The !Kung make their artifacts, on the whole, of durable material and take good care of them. They last for generations and move in a slow eddy among the people. The dealings are only between individuals, but they are numerous and provide occasion, perhaps more than any one other activity does, for visits which bring whole bands into polite encounter.

We gave cowrie shells as parting gifts in 1951 to the women in ≠Toma's band, the band which first sponsored us and with which we stayed and did our consistent work, wherever they were, on each expedition. We had carefully observed that there were no cowrie shells among !Kung ornaments before we gave them. Ours were bought from a New York shell dealer, perhaps to the confusion of future archaeologists. We gave to each woman enough for a short necklace, one large brown shell and twenty smaller grey ones. In 1952 there was hardly a cowrie shell to be found in Band 1. They appeared, not as whole necklaces, but in ones and twos in people's ornaments to the edges of the region.

The !Kung have not developed special objects to use as gifts. Nor have they invested ordinary objects with special gift significance. What they give each other are the common artifacts and materials of everyday life (see Appendix). Some objects are more highly valued than others as gifts, as one would expect. I gathered that relative scarcity of material was a factor and that objects were appreciated for their beauty, workmanship and appropriate size (a wide headband is better than a narrow one). People took an interest in remembering to whom an object had been given in the recent past, but the !Kung, who are present-oriented, do not place special value upon antiquity as such or systematically hold the distant past in mind.

The !Kung decorate their artifacts very little (they have developed music and dancing, not the plastic or pictorial arts), but they delight in ornaments with which to adorn themselves. The intrinsic beauty, the meaningfulness and the relief to the monotony of yellow-brown skin and grey-brown leather karosses are all valued. The most highly valued gifts are the traditional ornaments of ostrich egg-shell beads, especially the wide headbands and necklaces of five or six strings of beads reaching to the umbilicus. The creamy white is becoming to them. They like European beads also, of all colours, though white is definitely their preference. The only thing they do not like about beads, Gao Beard told us, is shortness of the string. After ornaments, the !Kung value implements or utensils, *//guashi* or dance rattles, which have required time and care to make, but perhaps more especially valued are metal objects and pots. The latter are obtained by trade. The !Kung do not trade among themselves. They consider the procedure undignified and avoid it because it is too likely to stir up bad feelings. They trade with the Bantu, however, in the settlements along the Botswana border. The !Kung offer well-tanned antelope hides and ostrich egg-shell beads. For these they obtain tobacco, beads, knives, axes, malleable metal for making arrow-points and spear-blades, and occasional files and chisels, fire-strikers and pots.

The odds are with the Bantu in the trading. Big, aggressive and determined to have what they want, they easily intimidate the Bushmen. Several !Kung informants said that they tried not to trade with Herero if it was possible to avoid it because, although the Tswana were hard bargainers, the Herero were worse. /Qui said he had been forced by a Herero whom he was afraid to anger to trade the shirt and pants we had given him as our parting gift in 1952 for a small enamel pan and a little cup. /Ti!kay had more gumption. A Herero at the beginning of a negotiation with him brought out a good-sized pile of tobacco but took from it only a pinch when it was time to pay. /Ti!kay picked up the object he was trading and ran off. ≠Toma said with amused exaggeration that 'a very good Herero, a respectable one, will give a handful of

tobacco for five cured steinbok skins. A bad Herero will give a pipeful (he showed the size of his fingernail) for three skins.' The Tswana values are a little better. A well-tanned gemsbok hide brings a pile of tobacco about 14 in. in diameter and about 4 in. high. The values vary. Some that were reported to us were three rooibok, duiker or steinbok skins for a good-sized knife, five strings of ostrich egg-shell beads for a spear.

The !Kung have become dependent on metal, especially knives, axes and arrow-points. They have been able to trade enough for every man to have these implements. They could, however, exist without them and do still use a few bone and wood arrow-points. Poison makes their arrows deadly, not penetrating power. The pots are Ovambo or Okavango pottery (the !Kung make no pottery themselves) or European ironware. The !Kung like to have a pot around to borrow sometimes; not everyone wants to carry one. They cook mostly in hot ashes. More for their novelty, I thought, than for their worth, the !Kung trade also for old oddments of cloth garments (they weave neither cloth nor mats), pieces of blankets, basins, etc. – things they do not really need – but which they like.

Tobacco they need 'to make the heart feel better'. Oddly enough for these passionate smokers, tobacco was not given so much emphasis in gift-giving as one might expect. They do make gifts of tobacco, but when anyone lights a pipe he passes it around anyway; all present drag smoke into their lungs until they almost faint and it has not mattered much who owned the tobacco.

Eland fat is a very highly valued gift. An eland provides so much fat that people can afford to be a little luxurious. They rub it on themselves and on their implements and they eat it. ≠Toma said that, when he had eland fat to give, he took shrewd note of certain objects he might like to have and gave their owners especially generous gifts of fat.

Real property and the sources of food and water are owned by the headman for the band and cannot be given away. However, meat, once it is distributed after the hunt, and veldkos, once it is gathered, become private property and may be given. Artifacts are privately owned by the individual man, woman or child, as outrightly owned if received as a gift as if made by the individual. They borrow and lend a great deal – in itself this is one of the ways they support each other and aid themselves in maintaining social solidarity – but this does not blur the clarity of ownership. Each object acquires some markings of its own from the maker and from usage. It is easy for the !Kung, with their highly developed powers of observation and visual memory, to keep track of the commonest objects, know the ownership and remember the history of the gifts.

As far as we know, no artifacts or materials are taboo as gifts. For instance, women may be given weapons, although it is taboo for women to touch them.

In quantity the gifts varied. One which /Ti!kay gave to ≠Toma was considered generous. It was a fine ostrich egg-shell headband, three ostrich egg-shells and a well-tanned duiker skin as soft as suede. Another generous gift consisted of a knife, a spear and a triple string of traded white European beads. Often gifts were less. The feelings persons have for each other, the degree of their past indebtedness, what they happen to possess and can give, determine the generosity of the gift.

The acquisition *per se* of the objects was seldom, I believe, of primary importance to most individuals in gift-giving – that is, if the objects were their own artifacts. As the !Kung come into more contact with Europeans – and this is already happening – they will feel sharply the lack of our things and will need and want more. It makes them feel inferior to be without clothes when they stand among strangers who are clothed. But in their own life and with their own artifacts they were comparatively free from material pressures. Except for food and water (important exceptions!) of which the Nyae Nyae !Kung have a sufficiency – but barely so, judging from the fact that all are thin though not emaciated – they all had what they needed or could make what they needed, for every man can and does make the things that men make and every woman the things that women make. No one was dependent upon acquiring objects by gift-giving.

They lived in a kind of material plenty because they adapted the tools of their living to materials which lay in abundance around them and which were free for anyone to take (wood, reeds, bone for weapons and implements, fibres for cordage, grass for shelters), or to materials which were at least sufficient for the needs of the population. The Nyae Nyae !Kung had hides enough for garments and bags; they kept extra hides when they needed them for new garments or when they wanted them for trade; otherwise they ate them. The !Kung could always use more ostrich egg-shells for

beads to wear or trade with, but, as it is, enough are found for every woman to have a dozen or more shells for water containers – all she can carry – and a goodly number of bead ornaments.

In their nomadic hunting-gathering life, travelling from one source of food to another through the seasons, always going back and forth between food and water, they carry their young children and their belongings. With plenty of most materials at hand to replace artifacts as required, the !Kung have not developed means of permanent storage and have not needed or wanted to encumber themselves with surpluses or duplicates. They do not even want to carry one of everything. They borrow what they do not own. With this ease, they have not hoarded, and the accumulation of objects has not become associated with status. Instead of keeping things, they use them to express generosity and friendly intent, and to put people under obligation to make return tokens of friendship. And, even more specifically, in my opinion, they mitigate jealousy and envy, to which they are prone, by passing things on to others. These are the objectives in gift-giving.

Except, as ≠Toma said, that it would be surprising to see a man give a present to a woman who was not related to him (and vice versa, I imagine), anyone may give to anyone. Degree or kind of kinship or affinity, having the joking relationship or lacking it, impose no requirements or restrictions, nor does status in the community in a conventional way. We did hear people say, however, that a headman may feel that he should lean well to the generous side in his giving, for his position as headman sets him out from the others a little and he wants whatever attention this attracts not to be envious. Someone remarked that this could keep a headman poor.

The times of giving are determined almost entirely by the individual's convenience. The !Kung do not know their birthdays or anniversaries and have no special days of the year which they mark by giving gifts. Gifts are required by convention on only three ceremonial occasions. The type or quantity of the gift is not patterned but it should be generous. The occasions are betrothals and weddings, when the parents exchange gifts and give to the young couple, and the ceremony of a baby's first haircut, when the *!gu!na*, the person for whom the baby is named, should give him a fine present.

Relatives give to young people with the idea of setting them up in life. Gao Beard gave a spear and a kaross to his FaSiHuSo, saying it was his duty to see that the boy got some things because among the boy's relatives he was the most able to do so. The boy's father was very old, he explained. People expect to wait a long time for young people to make return gifts.

The two rigid requirements in gift-giving are that one must not refuse a proffered gift and that one must give in return. Demi said that, even if he might prefer not to be obligated to someone, he would accept and prepare to make his return gift. If a gift were to be refused, he continued, the giver would be terribly angry. He would say, 'Something is very wrong here.' This could involve whole groups in tensions, bad words, taking sides – even 'a talk' might occur – just what the !Kung do not want. Demi said it does not happen: a !Kung never refuses a gift. And a !Kung does not fail to give in return. ≠Toma said that would be 'neglecting friendship'. A person would know that others thought him 'far-hearted' and 'this would worry him'.

In reciprocating one does not give the same object back again but something of comparable value. The interval of time between receiving and reciprocating varied from a few weeks to a few years. Propriety requires that there be no unseemly haste. The giving must not look like trading.

Incidentally, we were not included by the !Kung in their gift-giving patterns. They gave us a few things spontaneously which they thought we would enjoy – python meat, for instance – but did not feel obligated to reciprocate for every gift we gave them.

Asking for an original gift or asking that a return gift be made after due time has elapsed is within the rules of propriety. People prefer that others give in return without their asking, but ≠Toma says he does not hesitate to ask if a gift is long overdue. If a person wants a particular object he may ask for it. Asking is also a means by which people play upon each other's feelings. One can test a friendship in this way. One can give vent to jealousy or satisfy it by acquiring some object or making someone else uncomfortable. I thought that /Ti!kay (an intelligent man, but very touchy, unco-operative, and turned inward upon himself) used to ask for gifts to play with anger, arousing it for the sake of feeling it, as children do with fear, playing witches in the dark. His remarks one day indicated a mingling of feelings and purposes. He told us that one may ask for anything. He does, he said. He goes to a

person's fire and asks. (I could imagine him with his black glancing eyes sitting and asking!) He asks usually for only one or two things, but if a person has a lot he may ask for more. He said he is almost never refused but that if a man had only one pot and /Ti/kay asked for it the man might say, 'I am not refusing, but it is the only pot I have. If I get another you may come for this one. I am very sorry but this is the only pot I have.' /Ti/kay said this would not make him angry unless he were refused too many times. To be refused too many times makes a person very angry. But, said /Ti/kay, he himself did not tire of people asking him for gifts. Asking, he claimed, 'formed a love' between people. It means 'he still loves me, that is why he is asking.' At least it forms a something between people I thought.

I have stressed the mitigation of jealousy as the important value of gift-giving. /Kung informants stressed more the value of putting people under the obligation of expressing friendship by making the friendly gesture, even if it is only a token gesture. Demi said, 'The worst thing is not giving presents. If people do not like each other but one gives a gift and the other must accept, this brings a peace between them. We give to one another always. We give what we have. This is the way we live together.'

ABSENCE OF STEALING

One day, when I wanted to talk with a group of informants about what the /Kung considered to be wrong-doing, I began with /Ti/kay. He said promptly, 'Making crooked arrows and fighting', but could not think of anything else. Informants had previously said that not sharing food would be the worst thing they could think of. Others had mentioned that the breaking of the incest and menstruation taboos would be very wrong and that girls should not sit in immodest posture. No one seemed to think lying was very serious wrong-doing, and no one mentioned stealing. I finally asked directly and Gao Medicine said they had not thought to mention stealing because they did not steal.

We had heard of the episode of the man taking honey from a tree, honey which had been found and marked and was therefore owned by someone else, and being killed for it by the owner. That was the only episode of stealing that we discovered.

The /Kung stole nothing from us. Even when we went away on trips, leaving several bands of /Kung settled around our camp site, we left our supplies and equipment unlocked in the open or in our tents with confidence that nothing would be stolen. Things that we lost or forgot in the werf were returned to us, even two cigarettes in a crumpled package.

Stealing without being discovered is practically impossible in /Kung life because the /Kung know everybody's footprints and every object. Respect for ownership is strong. But, apart from that, /Ti/kay said, 'Stealing would cause nothing but trouble. It might cause fighting.'

CONCLUSION

During seventeen and a half months when I lived with the Nyae Nyae /Kung at Gautscha (with Bands 1 and 2 and many visitors – usually about sixty to seventy-five persons), I personally saw four flare-ups of discord and heard of three others which occurred in neighbouring bands during that period. All were resolved before they became serious quarrels. Of the seven, four were flare-ups of sexual jealousy. One was 'the talk' about Khuan//a's gift. Two were very minor disagreements about going somewhere, when Gao Beard's frustrated young wife hit him on the head with her digging-stick and when /Ti/kay gave his brother a shove for refusing to accompany him. None was about food. We considered, judging from that sample, that the /Kung managed very well to keep tensions from turning into hostility, to avoid arousing jealousy and ill will, to cohere and to achieve the comfort and security which they so desire in human relations.

APPENDIX

/Kung Bushman Artifacts

ANIMAL MATERIALS

HIDES

Made and used by men

- breech clouts
- oracle disks
- head-dresses (a few, worn in dances)
- caps (a few)

Made by men, used by women

- genital aprons
- small karosses worn around buttocks or for carrying babies

HIDES

Made by men, used by all
- karosses
- bags (sewn)
- bags of whole skins
- tobacco pouches of scrota
- bracelets, armbands, knee and ankle bands
- necklaces
- thongs

COCOONS

Made by men
- dance rattles

BONE

Made and used by men
- connecting parts for arrow-shafts
- arrow-points (few)
- dishes for mixing poison

Made and used by women
- fresh bone, pounded to powder, is used to clean and soften karosses

BIRDS' NESTS

Used by men and women
- tobacco pouches (nests of the penduline tit, *Anthoscopus minutus*)

SINEWS

Made and used by men
- bow-strings
- wrappings for arrow-shafts (the poison adheres well to sinew)
- binding for hafts
- snares

Made by men, used by all
- strings for *//guashi* (musical instrument)

Made by men or women
- fine rolled cord

HORN

Made and used by men
- horns fitted with leather covers for carrying arrow-points
- horns for cupping

Vegetable Materials

WOOD

Made and used by men
- bows
- spear handles
- axe or adze handles
- knife handles
- knife sheaths
- spring-hare hook handles
- fire-making sticks
- hatchet bow (musical instrument) (few)

Made by men, used by all
- digging sticks
- carrying sticks
- stirring sticks
- mortars and pestles
- bowls
- spoons
- fire scrapers
- *//guashi* (musical instrument)
- one-stringed violin
- *jani* toys
- beads

Made by women
- frames for scherms

FIBRES

(from *sansevieria*, aloe, *!noa* tree roots, other roots)

Rolled by men and women
- cord

CORDS

Made and used by men
- carrying-nets
- snares

Made and used by all
- ornaments

GRASS

Made by men and women
- stoppers for ostrich egg-shell water-containers
- ornaments

Made by women
- scherms

Made by children
- toys of grass folded and woven to suggest animals

Used by all for beds

REEDS

Made and used by men
- arrowshafts

BARK

Made and used by men
- quivers

Used by women
- strips for tying frames of scherms and for tying grass on scherms in times of high wind

Worn by all as head and chest bands

'TSI' NUTS

Used by all as beads

Used by boys as weights for *jani* toy

STONE

Made by men
- arrow-straighteners
- pipes

Used by men for
- sharpening blades
- hammers
- anvils
- weights to stretch nets
- marking graves

Used by women
- powdered red earth for ceremonial markings on face

Obtained by Trade with Bantu

MALLEABLE METAL AND WIRE

Made and used by men
- arrow-points
- spear-blades
- spring-hare hooks

Made by men, used by all
- pipes
- awls

Used by men
- knives
- axes or adzes
- chisels (few)
- files (few)
- fire-strikers (few)

Used by all
- beads – copper, glass, shell, etc.
- pots – Bantu clay, European iron
- pipes (few), European briar, old cartridge-shells
- receptacles (few)
- old scraps of garments and blankets
- oddments such as buttons for ornaments

Notes

1. This description of the !Kung in the Nyae Nyae region is based on observations made during our several expeditions between 1951 and 1959. Changes in the conditions under which the !Kung live are to be expected, and how these changes will affect the practice of the customs here described we do not know.

2. L. Marshall, '!Kung Bushman Band Organisation', *Africa*, vol. 30 (1960).

3. L. Marshall, 'Marriage among the !Kung Bushmen', *Africa*, vol. 29 (1959).

PART TWO

How People Live in Societies

'Personality and Psychotherapy' is something of a classic in the annals of psychological literature. John Dollard, a clinical psychologist of Freudian persuasion, collaborated with Neal Miller, an experimental animal psychologist, to produce a book that integrated these very different points of view. In it they discuss the learning of normal and abnormal human behaviour, regarding psychotherapy as the application of the same principles of learning to the unlearning of inadequate neurotic behaviour and the relearning of more appropriate 'normal' behaviour.

The excerpt chosen is concerned with the early socialisation process as it occurs in our own culture patterns. In it, four important areas of child training are analysed in a way that helps us to see how complex the process of socialisation is, what pitfalls await the unwary and how scientific principles can be applied to arrive at more rational child-rearing practices.

Ford and Beach's 'Patterns of Sexual Behaviour' is again something of a classic, in that data from 190 different societies, including our own, are used to explore the immense diversity of human sexual behaviour. To develop what they call a three-way perspective, the authors throw the cross-cultural material into relief by introducing two other types of approach, both biological. One takes the underlying physiological processes into account, and the other is comparative, considering man as part of the animal kingdom – albeit at its apex.

In the chapter presented here, the individual acquisition of sexual behaviour as a function of physiological changes and social pressures is compared with that of species at an earlier stage of evolution. Our preconceptions of what is 'natural' or 'unnatural' in sexual behaviour are likely to require careful reappraisal after considering this material. Perhaps its most immediate impact is to make us aware of the extent to which our own culture pattern engenders a sense of guilt about behaviour that is either commonplace in our own society, such as masturbation or homosexuality, or accepted elsewhere as entirely proper, such as an active and open sex life for children.

CRITICAL TRAINING SITUATIONS

John Dollard and Neal Miller

THE culture, of course, takes a position – a traditional position – on the various needs of the child. It has a design for the feeding situation, for cleanliness training, for sex training, for the treatment of anger responses in the child; and as the society imposes its will through the acts of the parents, the child reacts in its blind emotional way. Each one of the above-mentioned training situations can produce long-lasting effects on the character and habits of the individual and each is worth a brief discussion. We are by no means sure that these four are all of the dilemmas which can produce acute emotional conflicts, but we do know that each one of them has, in known cases, done so.

1. THE FEEDING SITUATION: CONFLICTS AND ATTITUDES

Much important learning takes place in reference to the hunger drive and the strong responses it excites. During the nursing period the child cannot 'comfort' itself. It cannot, so to say, tell itself 'It won't be long now', or 'Only twenty minutes till feeding time'. The hunger of the child is an urgent, incessant and timeless pressure which, obviously, produces the most intense activation. If the child is fed when hungry, it can learn that the one simple thing it can do to get results (i.e. cry) can make a difference in what happens. Learning to cry as a signal for food is one small unit in its control of the world. Such a trait could be the basis of a later tendency to be 'up and doing' when in trouble, of a belief that there is always a way out of a painful situation.

Apathy and apprehensiveness

If the child is not fed when it is crying, but is instead left to 'cry itself out', it can, similarly, learn that there is nothing it can do at that time to change the painful circumstances. Such training may

also lay the basis for the habit of apathy and not 'trying something else' when in trouble. In a second case, when drive is allowed to mount, the child can also learn that being a little bit hungry is followed by being very painfully hungry. When the child is then fed, only its most violent responses are reinforced. In this case the child can learn to fear being very hungry when it is only slightly hungry and to make the frightened response appropriate to severe hunger when only mild hunger exists. It is thus learning to 'over-react', to be apprehensive of evil even when the circumstances of life seem calm. This learning occurs through the behaviour mechanism of anticipation.

Sociability and 'love'

On the other hand, probably the feeding experience can be the occasion for the child to learn to like to be with others; that is, it can establish the basis of sociability. When the hungry infant is fed, some of the wonderful relaxation responses which it experiences can be conditioned to the stimuli of those persons who are caring for the child. Thereafter the mere appearance of the mother can produce a momentary feeling of well-being. The child will learn to stop crying at the sound of her footstep, the rustle of her dress or the sound of the tap water which is warming its bottle. These experiences have an intense emotional quality which is often attached thereafter to the word 'mother' as the source of all beneficence.

Likewise, if the child is properly held, cared for and played with, the blessed relaxing quality of these experiences also will attach to those who care for it. Since the mother or caretaker stands at the very head of the parade of persons who become 'society' for the child, it is quite important that she evoke such benign and positive responses in the child.

Lack of social feeling

The reverse of all this can also take place if the child is stuffed when it is not hungry. If its food rewards are in various ways cut down and spoiled, it may not care much whether 'the others' are there or not. It may tend to be 'low in social feeling'. If the child is actually punished for crying when it is hungry, as by being slapped, a true hunger-anxiety conflict will be created. Though this may be rare, it does undoubtedly happen, especially when the child is overactive as a result of gastric upset and so is able to provoke anger in the ill-disciplined parent.

One origin of fear of being alone

The child can learn another dangerous habit in this period. It can learn to fear being alone. Teaching a child to fear being alone is easy to do and is often done inadvertently. Let the child get very hungry when it is alone, let it cry and not be heard or attended to, but let the quantity of stimulation in its body from hunger and from crying continue to rise. When the child is finally fed, these very strong terminal responses are reinforced and can be attached to all the stimuli which were present during the period of its intense hunger. These responses can produce stimuli of drive-like strength. Similarly responses which produce strong drives can be attached to the darkness, to the immobility of objects, to quietness, to absence of parental stimuli. Once the child has inadvertently learned to 'fear' darkness and quietness and immobility, it will also learn to escape from the darkness into the light, from the quietness into noise and from immobility into the presence of others.

This escape may be perceived by the parents as an additional nuisance when they are expecting their hours of relaxation; the child insists on being with them even though 'there is nothing wrong with it'. They may then take punitive measures, forcing the child back into the dark or the quiet and creating a true conflict between fear of darkness and the newly learned fear of the irate parent. This must indeed be a very common conflict, since fear of quietness and darkness are not innate in children and yet are frequently seen.

If this fear persists into adult life, it can be an element in the character of a person who is compulsively driven to social contacts, who cannot tolerate being alone. Compulsive sociability may also involve a sacrifice in creativeness, since in order to be creative the individual must be able to tolerate a certain amount of loneliness.

Weaning

In the case of weaning also, severe traumatic circumstances may arise. If the child is suddenly changed from one type of food or mode of feeding to another, it may go on a hunger strike which the parents obstinately oppose, saying 'It will eat when it gets

hungry enough.' Indeed it will, but in the meantime it may have learned some of the fears or the apathy already listed. If parents punish the child for its refusal to eat the new food, a genuine conflict is created which in turn will have its consequences. There seems hardly anything valuable that an infant can learn by punishment under such circumstances, and parents should take the greatest pains to avoid this.

Colic and recurring hunger

The child with colic is also a sore trial to itself and its parents. One of the simplest circumstances producing 'colic' is that the infant has eaten too much and must regurgitate some of the food or the gases which its digestion produces. Once it has been laboriously walked or patted into parting with food or gas, it may be hungry again. Unimaginative parents, not understanding that hunger has innocently recurred, will fail to feed the child. If the mother does feed it, the child may overeat again and the cycle of gastric tension, vomiting and hunger may recur. However, until an infant learns to make its gastric distress anticipatory and thus to check itself while eating, there is no way of avoiding these circumstances. The sequence overeating, gastric distress, vomiting and recurring hunger seems more likely to occur with children fed on schedule, since they will get much hungrier while waiting for the scheduled moment of feeding and are more likely to overeat.

If parents lose their tempers and punish the young child at any phase of this awkward kind of learning to eat just the right amount, severe conflict concerning feeding may ensue. If the infant is punished before it is burped, the result may be that it has anxiety attached to burping and regurgitating, and it is thus condemned to bear gastric tension. If it is punished after regurgitating when it is again hungry, anxiety responses will be attached to hunger stimuli. Under these conditions punishment cannot teach the infant anything that will help it along its road of development. Nevertheless this unavoidable circumstance of colic is one to test the character of the most devoted parents.

The foregoing discussion is by no means a check list of all the things that a child can learn during the first year or so of life. For example, in learning to crawl and walk the child is also learning to fear bumps. It learns not to poke its head under the table and then suddenly try to stand erect. It is learning a few words and common commands. . . . Various specialists in pædiatrics such as Spock have described the behaviour problems that are most frequent among young children in the home.

Secret learning of early years

What we have attempted to do here is to show that the seemingly innocuous feeding situation can be fraught with important emotional consequences. Outsiders who cannot know what is going on in a home may see no reason to suppose that the infant is learning anything at all. Yet observant insiders may see the child becoming apathetic, apprehensive, learning to fear the dark, on the one hand, or becoming loving, sociable and confident on the other. It is this secret learning of the early years which must be made the object of scientific research. We are firmly of the opinion that anything that can be sensed can be scaled and thus that apathy, sociability and fear can be scientifically treated if we but trouble to study the child in the home – where these habits are being learned.

Early conflicts unlabelled, therefore unconscious

The young child does not notice or label the experiences which it is having at this time. It cannot give a description of character traits acquired during the first year of life nor yet of its hardships, fears or deep satisfactions. What was not verbalised at the time cannot well be reported later. An important piece of history is lost and cannot be elicited by questionnaire or interview. Nevertheless, the behavioural record survives. The responses learned occur and may indeed recur in analogous situations throughout life. They are elicited by unlabelled cues and are mutely interwoven into the fabric of conscious life. The fact that different children learn different things during this period undoubtedly accounts for some of the variability between children which is often attributed to innate factors.

2. CLEANLINESS TRAINING CAN CREATE CONFLICTS

If the child has come safely and trustfully through the early feeding and weaning experience, it may learn for the first time in its cleanliness training that the culture patterns lying in wait for it have an ugly, compulsive aspect. No child may avoid this

training. The demands of the training system are absolute and do not take account of individual differences in learning ability. The child must master cleanliness training or forfeit its place in the ranks of socially acceptable persons. Freud describes the culture's task as building within the personality of the child the psychic dams of loathing and disgust for urine and faeces and particularly for the latter. The attempt to construct these inward barriers immediately puts the child in a conflict situation.

Observation of children within the home indicates that children begin with the same naïve interest in their faeces and urine that they have in the other parts and products of their bodies. Development of the ability to grasp and finger objects makes it possible for the young child to handle and play with faecal material. The morning will arrive in every nursery when the astonished parents will observe their beloved child smearing faeces over his person, his hair and his immediate environment with gurgling abandon. This may be the first occasion for sharp, punishing exhortations, for angry dousing, for the awakening of anxiety in connection with faecal materials. On pain of losing the parents' love and so exposing itself to the high drives and tensions which occur when they do not support it, and on further pain of immediate punishment, the child must learn to attach anxiety to all the cues produced by excretory materials – to their sight, smell and touch. It must learn to deposit the faeces and urine only in a prescribed and secret place and to clean its body. It must later learn to suppress unnecessary verbal reference to these matters, so that, except for joking references, this subject-matter is closed out and excluded from social reference for life.

Difficulty of cleanliness learning

Cleanliness training is difficult because culture must work a reversal of a strong innate connection between a cue and a response. The swelling bladder or bowel produces a strong drive stimulus which at a certain strength releases the urethral sphincter or touches off the evulsion response in the anus. To meet cultural demands this sequence must be rearranged. The connection between bowel stimulus and the evulsion response must be weakened. The child must learn to suppress the evulsion response to the bowel drive-stimulus alone. It must then insert other responses in the sequence. At first it must learn to call to the parents. It must later learn to insert walking, unbuttoning and sitting on the toilet chair while it is still suppressing the urgent evulsion response. Only to a new pattern of cues – the bowel stimulus, the cues of the proper room, the sense of freedom of clothes, the pressure of the toilet seat on the child's thighs – may the evulsion response occur without anxiety.

In short, this response occurs not only to the pressure of the primary drive involved but also to the complex stimulus pattern just named. If one can once get the child to order the responses correctly, the strong tension reduction produced by defaecation will reinforce the responses to the pattern of cues enumerated. The real problem, therefore, is getting the child to suppress the naïve evulsion response and to insert a considerable series of responses into the sequence before evulsion.

We do not revel in the details of this analysis but offer the detailed analysis because we believe it is impossible to understand the difficulty of the learning involved unless one sees all the new units which must be learned. For instance, buttoning and unbuttoning is a difficult habit for small children to learn and may hold up the perfect learning of the sequence for some time. The child, however, is not really trained until it can carry out the whole sequence by itself.

Learning without verbal aids

The difficulties which produce conflict in this learning arise chiefly from the fact that the child must accomplish it in a period of life when it has to learn mainly without verbal aids, that is, by trial and error. Learning cleanliness control by trial and error is a slow and vexing business. The child must learn to wake up in order to go to the toilet, though sleep seems good. It must learn to stop its play even when social excitement is strong. It must learn to discriminate between the different rooms of the house – all this by crude trial and error. In this case, 'trial' means urinating or defaecating in an inappropriate place, and 'error' means being punished for the act so that anxiety responses are attached to the cues of this place. In the trial-and-error situation this must be repeated for each inappropriate place – bed, living-room, dining-room, kitchen, 'outside'.

The function of this training is to attach anxiety responses to the defaecation drive so that they win out over the immediate

evulsion response. These anxiety responses also motivate and cue off the next responses in the series, such as calling to the parents, running to the bathroom, unbuttoning the clothes and the like. When accomplished by trial-and-error means, this training necessarily takes considerable time, perhaps several years in all, in which child and parent are under severe pressure.

Strong emotions aroused in cleanliness training

Learning cleanliness is no mere behavioural routine. It arouses strong emotions – perhaps as strong as are ever evoked in the child again. Anger, defiance, stubbornness and fear all appear in the course of such training. Fear may generalise to the toilet itself and excite avoidance responses in the very place where the child is expected to 'go'. Unable to discriminate between the safe and the unsafe place, the child may try 'not to defaecate at all'. This behaviour is perfectly automatic, but it may seem wilful to the parents, and they may particularly resent the final loss of control after the protracted attempt to inhibit defaecation. Once hit on, this response would be strongly reinforced and tend to become habitual, since the drive reduction after prolonged withholding would be much more intense than after a normal period of withholding. When 'losing control', instead of deliberately relaxing, is strongly rewarded, the habit of 'losing control' should become anticipatory and thus prolong the problem of cleanliness training. In other words, great strictness at early ages may block rather than advance the child in his cleanliness learning.

Learning to escape from sight of parents

The child may become, from the parents' standpoint, furtive by the following means: When it is punished for a cleanliness error by the parent, anxiety is attached to the sights and sounds produced by that parent. In order to escape that anxiety the child may attempt to escape from the parental presence and attempt to keep to a minimum the amount of time it spends near the parent. This state of affairs has the disadvantage that the child is escaping from one of its natural teachers. It may learn to speak less well than it might because it simply does not remain near those people who could teach it to speak. Infliction of punishment may also arouse anger towards the inflicting agent. The child may attempt struggling with the parents, biting them or slapping at them and, in turn, be punished for this behaviour. Thus an anger-anxiety conflict is learned.

Excessive conformity and guilt

Again, the child may get the impression that it is pursued by an all-seeing, punishing guardian and may try making as few responses as possible – and certainly not innovating any novel responses. Its conclusion on the basis of punishments received may be that unless a response is known to be correct it should not be risked. Thus may be laid the characterological basis of the excessively timid, conforming individual. Similarly the child may not be able to discriminate between parental loathing for its excreta and loathing for the whole child himself. If the child learns to adopt these reactions, feelings of unworthiness, insignificance and hopeless sinfulness will be created – feelings which sometimes so mysteriously reappear in the psychotic manifestations of guilt.

Advantages of verbal aids

From this discussion it will be clear that the trial-and-error method of early training, with its many punishments, has much more risk attached than training carried on at a time when the child can be verbally aided to hit on the right sequence of responses in a few early trials. Once the child has acquired the words 'living-room', 'kitchen', 'bedroom' and 'outside', a single punishment trial, if properly conducted, can attach anxiety to all these cues at the same time and so spare the brutal repetition of punishment. If the child has already learned to call for help when it needs help, it can much more easily learn to call for aid when it needs to defaecate. If it has learned to stop various activities to the word 'stop', it is much easier to get it to check the evulsion response when this is occurring to its innate stimulus. If certain promises of the parents already have reward value attached to them, the child can be aided to make the right responses by being promised simple rewards. If the child already attaches anxiety to certain instructions of the parents, these instructions can have some of the same effect as repeated, direct punishments.

In this case also the reinforcement of the act of defaecation itself will fix the correct series of responses into place. This will happen whether the course of the training has been stormy or smooth. However, in the case of the smooth, verbally aided learning there

is much less danger of arousing furious anger or of creating maladaptive habits such as retention of faeces and loss of control. Extremely strong anxiety reactions do not occur and feelings of excessive worthlessness are less likely. The end result is the same so far as mere cleanliness training is concerned. The difference lies in the fact that the later, verbally aided method of getting out the response has much less risk of violent side reactions and character distortions.

Freud's superego

The foregoing analysis employs the thoughts and sentences of Freud reworked from the standpoint of behaviour theory. The course of cleanliness training is unlabelled and unconscious. Any one of us may have been through a stormy period of this kind and yet have no recollection of it. The results may show themselves in our symptoms, our most deeply embedded 'character' traits, in our dreams, in our intuitive presuppositions about life, but they will not show themselves in our verbal behaviour. The record of this training will be found in no man's autobiography, and yet the fate of the man may be deeply influenced and coloured by it.

The first broad strands of what Freud calls the superego are laid down at this time. Anxiety reactions, never labelled, are attached to stimuli, also unlabelled. When these stimuli recur later the anxiety reactions automatically recur. The resulting effect Freud has called the 'superego' or unconscious conscience. When unconscious guilt reactions are severe, the personality is suffused with terror. It is hard to say whether a morbid conscience is a worse enemy of life than a disease like cancer, but some comparison of this kind is required to emphasise the shock produced in the witness when he sees a psychotic person being tortured by such a conscience. Enough is known now to convince us that we should make the humble-seeming matter of cleanliness training the subject of serious research.

3. CONFLICTS PRODUCED BY EARLY SEX TRAINING

Sex-anxiety conflicts seem frequently to be involved in neuroses arising in civilian life. The recurrent appearance of sex as a conflict element does not seem to be due to the fact that sex is the strongest of human drives. At their highest levels, pain, hunger and fatigue certainly outrank it. Many strong secondary drives such as anxiety, ambition and pride can also be stronger than sex. Sex seems to be so frequently implicated because it is the most severely attacked and inhibited of primary drives. Even though relatively weaker, sex can exert a strong pressure which produces great activation in the organism and great misery if blocked for long periods. In no other case is the individual required to wait so many years while patiently bearing the goading drive.

Source of first sex conflict: the masturbation taboo

Erection of the penis can be observed in male infants as a reflexive response to interrupted feeding or to urethral drive pressure. At the age of a year the child is able to grasp an object quite perfectly. The sensitivity of the genital and the ability to prehend make masturbation possible. It seems likely also that there is some kind of reward associated with masturbating. On the basis of his observations, Kinsey believes that small boys acquire the capacity for orgasm long before they become able to ejaculate; similarly an experiment by Sheffield, Wulff and Backer demonstrates that sexual responses short of ejaculation can serve to reinforce learning in the albino rat. It is certainly a fact that, if unchecked, children do learn to masturbate and that they sometimes obstinately persist even when quite severe sanctions are applied.

The sight of a child masturbating evokes intense anxiety in the adults of our culture and they promptly apply sanctions, ranging from persistently removing or jerking the child's hand away from its genital to slapping and spanking it. The result is to set up in the child the same sex-anxiety conflict which the adults have. As in other cases, masturbatory conflicts established in the first years of life are invariably unconscious. A vague negative feeling, a tendency to withdraw, an unease is established at the act, sight or thought of masturbatory behaviour. These conflicts differ for different individuals in many ways and for many reasons. Some individuals may be caught in the act more often than others; some may be punished more severely than others; some may have stronger innate sex drive than others. Some may have had more time to learn the habit before being caught and punished and may thus have a stronger appetite for this behaviour than other persons. Some may, so to say, scare easier than others because they already have strong anxieties established in the cleanliness-training situa-

tion. Such anxieties generalise easily from urethral to the genital stimuli. Often both are called 'nasty' and the cue produced by the common verbal response helps to mediate generalisation of fear. In this case it is easy to train the individual out of the masturbatory habit, since the fear does not have to be learned but only generalised to the sex stimuli.

Parents don't notice effects of taboo

The imposition of a masturbation taboo can have important effects on the child's life. There may be immediate and direct changes in behaviour of the kind described as a 'bed phobia'. When behaviour changes occur it seems quite surprising that parents do not notice them as results of conflict over masturbation. The fact that they do not so notice is, however, easily explained. Intimate as their contact is with the child, they may yet be very poor observers of cause and effect. Most of the young child's emerging life is mysterious to parents anyway. They may further have particular avoidances against noticing matters and connections which arise in the sexual sphere. Likely, they believe themselves to have been sexless in childhood and can do no less than believe the same in respect to their children. Whether correctly evaluated by parents or not, the masturbatory taboo is the first of the important sex taboos, and it sets up a sex-anxiety conflict in each of us.

Sex typing of personality

The sexual development of the child cannot be understood without understanding the forceful training in sex typing which it receives. The unspecialised or less specialised human being, the infant, is identified as boy or girl and its relationship with others is defined in terms of sex type. Sex typing is a strictly conventional arrangement that varies from society to society. Our own society is strongly organised around sex specialisation of personality. This begins with male and female names, clothes, play patterns, toys, and continues throughout life by defining specialised sex roles for man and woman. The ultimate love object of the child is defined as a member of the opposite sex. The nascent sexual reactions of the child are directed towards stimuli of the opposite sex. The child is led to expect eventual sex rewards from persons of the opposite sex.

The taboo on homosexuality

Training in sex typing has the indirect effect of imposing a vigorous taboo on homosexuality. Homosexual objects are not presented, are treated by neglect or, if need be, vigorously condemned. The errors children make while learning sex typing are the source of much amusement to adults. The little girl declares she is going 'to marry mommie' when she grows up or the little boy states he will marry his admired older brother. Children are carefully corrected and trained into making the appropriate distinctions. Furthermore, it seems probable that parents, already sex-typed, help to develop this turning towards the opposite sex by themselves 'favouring' the child of the opposite sex.

Students of sexual abnormalities have suspected that the failure to define sharply the sex type is a factor in producing perverse sex adjustment. Thus, if a boy child were ardently desired, the parents might fail to impose sharp feminine sex typing on the girl who actually arrived. Or, in the opposite case, a mother who prefers her son to remain her 'baby' may make him effeminate when she should be emphasising his masculine character. Such inversions of social sex typing cannot directly produce a sexual perversion since sex responses must be attached to same-sex cues before a perverse sex appetite can exist; but they might tend to confuse the child about what its socially expected sex goals were and thus contribute to deviation.

After sex typing has been imposed and well learned, the child is in about this net position: masturbation has been tabooed, and it cannot give itself sex rewards by this means; sex behaviour between siblings has been suppressed; on the other hand, a new channel, though a long one, has apparently been opened through the fact of sex typing. The child is vaguely led to expect something rewarding in the general direction of the opposite sex. These two circumstances set up the situation of the Oedipus complex.

How fear is attached to heterosexual approach responses

The anxiety which adolescents, and often adults, show at the prospect of heterosexual contact must be explained. It does not arise by chance. It arises rather in the family situation which is the child's most important early learning situation. The first definition of sexual responses is learned in relation to parents and

siblings and only later transferred to others. Freud calls this the Oedipus situation.

We will illustrate from the case of the boy child, where the matter seems to be clear, and rehearse and paraphrase the familiar facts discovered by Freud. The boy child turns to his mother in fact or thought in the hope of getting sex rewards when he can no longer get them by himself. He expects sex rewards partly by generalisation of expectation of reward – that is, by analogy to the many rewards the mother has already given him – and partly from the fact that by sex typing he has learned to expect sex rewards from a woman and his mother is the woman at hand. Doubtless some of the anxiety already learned in connection with masturbation generalises to the sex impulse when it begins to show itself towards family women.

A new source of anxiety appears, however; that is, fear of the father. The five-year-old boy knows his father is the head of the house, the symbolic source of punishments and discipline. He also knows that his father is the husband of his mother and has some unique relationship to her. This rivalry of the father does not exist merely in the boy's mind. It is often made very concrete in the father's behaviour. The father may complain that the little boy sleeps in the mother's bed when he is already 'too old' for such behaviour. The father may object to the fact that the child or children sop up so much of the mother's time and leave so little to him. The father may impose certain restrictions about entering the parents' room which leave the child with a mystery on his hands. Whenever the male child makes emotional demands on the mother, the father may become more critical of him in other and more general respects, saying that the boy talks too much, that he does not work enough, and so forth. If the boy reacts with fear towards his father as a rival, it is because the father, consciously or unconsciously, is acting in a way that seems fearsome and rivalrous. The child is usually unable to discriminate between opposition on ground of sexual leanings and that evoked by its other claims on the mother. The whole thing may be played out as a kind of dumb show. The heterosexual strivings of the boy towards the mother may be behaviourally real and active but not labelled in the boy's mind. On the other hand, the opposition of the father, though active and effective, may be oblique and unconscious.

Often the mother herself rejects the claims of the boy. She has anxiety at any overtly sexual responses from the child, stops fondling him, and may suddenly and inexplicably change from being loving and approving to being horrified, disgusted and disapproving.

In this case there is less need for the father to be harsh and hostile. But if the mother does not reject and does not clearly show her separate loyalty and adherence to the father, a great burden is placed upon him to maintain his control of his wife. The mother, for example, may use the seeming need of the child as a way of escaping from her husband and from the sexual conflicts which she has in regard to him. She may favour and cozen the son while avoiding her husband, and unconsciously this may seem to the father like a genuine kind of preferment. The father may then react by very actively arousing the boy's fears.

Specific genital anxiety

If the boy's motives are sexual, the increased threat from the father produces anxiety which is directly attached to the sexual motives and interpreted as a sexual threat. This is one way in which castration anxiety may become an important factor in the boy's life even though the father never threatens castration in so many words. The boy has learned that the punishment often fits the crime.

There are other and less ghostly sources of the castration threat. Very often it has been specifically associated with the masturbation taboo – i.e. that if the boy plays with his penis, the penis will be cut off. The threat may appear in the fables of childhood which are told so eagerly. One of the authors as a six-year-old boy was permitted to participate in an after-dark session of older boys. They were telling the tale of how Bill Smith, a prominent citizen of the town, had come home and surprised his wife in bed with her lover. Smith thereupon pulled out a spring-bladed jack-knife (demonstration of length and viciousness of same by boy telling the story) and proceeded to unman the lover. Such a story does not remain, however, as a mere 'fable'. It is taken to heart and has the effect of teaching straight-out castration fear to sex motives.

The castration idea may occur in still another way; that is, as an inference from the lack of penis in the girl. The parents do not explain the different nature of the girl's genital. The uninstructed

boy may assume that the girl once had an external genital but has been deprived of it, perhaps as a punishment. There is no doubt that this inference is often made. The authors have repeatedly heard it in those in-family situations where children are first questioning their elders about sexual matters. It is further surprising in the history of adults how often the idea of bodily damage occurs in relation to sex 'sins'. Castration fear has been shown clinically to be connected with fears of bodily damage, especially in the cases of heart and brain, to aversion to crippled people, and to avoidance of women in their genital aspect. Castration fear is frequently escaped by approaching the bachelor girl (who has no husband or father at her side) or by resource to women of lower class or racial status (whose normal protectors are not allowed to function).

In any case, and engendered by whatever of these several means or combination of them, the sex conflict takes a new twist when it is worked out within the family. Anxiety which was once attached only to the masturbation impulse is now attached to the heterosexual approach situation. If this anxiety is made very strong it can produce a certain relief in the intensity of the conflict. This is the so called 'resolution' of the Oedipus complex. When anxiety is greatly dominant over approach tendencies, the conflicted individual stays far from his goal and but few of the acquired elements in the sexual appetite are aroused. Thus, that part of the intensity of the conflict which is produced by appetitive sex reactions is missing, and the conflict is therefore lessened. However, this conflict should and does recur when the individual is placed near his goal object and cannot easily escape, as frequently happens in adolescence. Then again the full strength of the sex reactions is pitted against the terror of sexual injury. Marriage evidently seems to some adults a similar situation – that of being held close to a feared goal – and they make the blind escape responses which would be expected.

Heterosexual conflict not labelled

If the prior intimidation of the person has been very great, and if the mother's stand is correct, much less fear need be imposed by the father. If sex appetite is weak rather than strong, there is much less pressure from the child's side and less anxiety need be imposed to counteract it.

All these events are but poorly labelled at the time they occur. The culture is niggardly about giving names to sexual organs, sexual feelings or the fears attached to them. The child is therefore not able to make a logical case for itself and, so to say, 'put it up to the parents'. Furthermore, repression sets in in two ways: Children are frequently forbidden to talk to others about their sexual reactions. Such sentences or thoughts as do occur tend to make the conflict keener both by arousing sex appetites and by cueing off the anxiety attached to them. The child is pained when it tries to think about sexual things and relieved when it stops. The result is repression. This repression has one unfortunate consequence for science. When the individual is later interviewed he is not able, promptly and freely, to give account of these matters. The renaming and mental re-establishment of these bygone events can thereafter only be made through the weary work of psychotherapy.

Science is not the only loser. The individual himself has lost his opportunity to use higher mental activities in solving the conflicts involving sex and authority. There are many ways in which the person can be victimised. A sexual perversion may lurk behind the blank surface of repression. The individual may never again be able cheerfully and amiably to accept a measure of authority exerted over him. Acute anxiety may be attached to his heterosexual impulses and when the time comes that society expects, almost requires, that he marry, he may be unable to do so. Even though he is able to get over the line into marriage, he may find the years of his marital life haunted and poisoned by constant, unconscious anxiety. In this case, the individual has automatically generalised to all women the anxiety proper only to the incest situation. He has failed to discriminate, as a free mental life would enable him to do, between the tabooed sexual feelings and objects of childhood and the relative freedom permitted to adults. To every authoritarian figure in his life he generalises the intense anxiety that he once experienced when attempting to rival his father in the sexual field. Only when higher mental processes are restored can the individual make those discriminations which allow him to proceed freely and constructively with his life as an adult.

4. ANGER-ANXIETY CONFLICTS

At this point we are more interested in the connection between angry emotions and fear than we are in the problem of how angry feelings are aroused in the child. We assume, however, as before that anger responses are produced by the innumerable and unavoidable frustration situations of child life. In the frustration situation, new and strong responses are tried out. Some of these have the effect of inflicting pain on other people. Society takes a special stand towards such anger responses, generally inhibiting them and allowing them rein only in a few circumstances (self-defence, war, etc.). Many of these attack, or 'put through the act', responses produce strong stimuli, and these we recognise as the emotion of anger. Lift the veil of repression covering the childhood mental life of a neurotic person and you come at once upon the smoking responses of anger.

Patriarchal code on child's anger

Parents intuitively resent and fear the anger and rage of a child, and they have the strong support of the culture in suppressing its anger. Direct punishment is probably used much more frequently when the child is angry and aggressive than in any other circumstance. More or less without regard to what the child is angry about, fear is attached to the stimuli of anger. The virtuous chastisement of the rebellious child is an age-old feature of our patriarchal culture. According to the old Connecticut Blue Laws, a father could kill a disobedient son. Even though this code was never exercised in this extreme in recent times, it shows the complete freedom to punish which was once culturally allowed parents. As the domestic representative of the patriarch in his absence, the mother is free to punish children 'in their own interest'.

How fear is attached to anger cues

We have already noted the situation of early cleanliness training as one tending to produce angry confusion in the small child. At earliest ages the cultural practice seems to be that of extinguishing anger rather than punishing it; that is, the child is segregated, left to 'cry and thresh it out'. However, parents' motivation to teach the child cleanliness training is so strong that they frequently also use punishment, especially in the case of what they interpret as stubborn or defiant behaviour. Anxiety responses therefore become attached not only to the cues produced by the forbidden situation but also to the cues produced by the emotional responses which the child is making at the time. It is this latter connection which creates the inner mental or emotional conflict. After this learning has occurred, the first cues produced by angry emotions may set off anxiety responses which 'outcompete' the angry emotional responses themselves. The person can thus be made helpless to use his anger even in those situations where culture does permit it. He is viewed as abnormally meek or long-suffering. Robbing a person of his anger completely may be a dangerous thing since some capacity for anger seems to be needed in the affirmative personality.

Other frustrations producing anger

The same state of affairs can prevail and be additionally reinforced as a result of the frustrations occurring in the sex-training situation. If the child is punished for masturbating it may react with the response of anger. The parent may not notice the provocative circumstance but see only that the child has become mysteriously 'naughty'. Its naughtiness may be punished and the connection between anger and fear be strengthened.

Parental rejection or desertion may likewise produce anger in the child. If the child feels secure only when the parents are present, it may react with fear when the parents leave or when they threaten to leave again. When the parents return, the child may make excessive claims, want unusual favours, 'be clingy'. To these demanding and possessive gestures on the part of the child the parent may react with unintelligent punishment, thus again teaching the child to fear.

The new tasks involved in growing up impose many frustrations on the child. Giving up long-standing privileges may arouse rage. Being forced to try out new responses, such as putting on its own clothes or tying its own shoe-laces, can anger the child. If it screams, lunges, slaps at the parent in these circumstances, punishment is the almost inevitable answer, and the connection between anger and fear is additionally strengthened.

Sibling rivalry

Rivalry between siblings is a constant incitement to anger, and such rivalry occurs in every household, without exception, where there are siblings of younger ages. The occasions for rivalry seem innumerable. Siblings may compete for evidences of parental love. If the parent disappoints a child, that child may 'take it out' on the luckier brother or sister. Younger children may anger older ones by being allowed to assume too quickly privileges which the older have long waited and worked for. Older children may tease and torment younger ones in retaliation. Sometimes the younger child is resented merely for existing and for having displaced the older one and alloyed its satisfaction in being the unique child.

The younger children may enjoy privileges which the older have been forced to abandon and thus create some degree of unconscious resentment. Younger children may tyrannise over older ones by too freely playing with or even destroying their toys and precious objects. Parents should intervene and prevent such behaviour, but often they do not, and the older child revenges himself in roundabout ways. Younger children may resent the privileges enjoyed by the older and attempt to punish older siblings for their greater freedom. These angry displays result in punishment of the one or the other child by the parents – and sometimes of both. The younger children tend to 'catch it' more from the older, and the older children more from the parents. Though parents may mitigate these angry relationships between siblings by just rules which are honestly enforced, there seems no way to take all the hostile strain out of such relations.

Mental limitations

Small children confront an unintelligible world. Many of their frustrations result from this fact. They do not have the mental units to be patient and foresightful. They do not know how to comfort themselves while waiting. They cannot live in the light of a plan which promises to control the future. Since so much is frustrating to them that is later bearable, they are especially prone to anger. They want to know, 'Why isn't the circus here today?', 'Why do I have to wait till my birthday to get a present?, 'Why does Daddy have to go to work just when it's so much fun to play with him?' Living in the present and being unable to reassure themselves about the future, young children resort to anger at these inevitable frustrations. Adults experience the hostile or destructive behaviour of young children as a nuisance, do not understand its inevitability, and frequently punish aggressive responses.

Devious aggression

If anger must be abandoned as a response in a frustrating situation, other responses will be tried out such as pleading for what one cannot take by force or submitting to frustrations which can only be worsened through opposition. Devious forms of aggression are particularly likely to occur in this case. The individual can be punished for direct anger responses but it is much harder to catch him at roundabout aggression. He may learn to lie in wait and take revenge by hastening and sharpening punishment which his opponent has invoked in some other way. Gossip, deceit, creating dangerous confusion about agreements and life relationships may all be indirect modes of angry reaction.

Anger conflict unlabelled

As in the case of sex-anxiety conflicts, the anger-anxiety conflict is likely to be poorly labelled. Verbal skills are at a low level when much of this training is going on. Repression of the language describing anger-anxiety conflicts may occur because conflict is thus, momentarily at least, reduced. As a result the individual cannot, in later life, be selectively angry, showing anger in just those social situations in which it is permitted and rejecting anger where it is not.

The overinhibited person

Inhibition of anger may occur in two different degrees. The overt, or some of the overt, responses of direct aggression may be inhibited. Some such inhibitions must occur if a child is to live in our culture. The process may, however, go farther and the emotion of anger itself be throttled. If the response-produced drives of anger evoke intense fear, the individual may be incapable of a normal life. The victim loses the core of an affirmative personality. He may be unable to compete as is demanded by our society in school or business spheres. He may be additionally shamed because he cannot bring himself to fight. He may depend unduly on others, waiting for them to give him what is everyone's right to take. Such a child cannot be a self-maintaining person because

he cannot produce any anger responses at all, let alone those which are 'legitimate and proper'.

Since many outlets for anger are permitted to adults which are not permitted to children, the person who is overtrained to inhibit anger may seem childish in that he is still following the age-graded code of childhood and is unable to embrace the freer standards of adulthood. One of the chief tasks of psychotherapy, in the case of unduly inhibited persons, is to enable them to name and describe their angry feelings so that they may extinguish undue fear and begin to learn a proportionate self-assertiveness.

FRUSTRATED MOBILITY ASPIRATIONS PRODUCE AGGRESSION

There is little doubt that adults can be in conflict concerning their mobility strivings and that these conflicts can lead to pathological results in behaviour. The conflict could be described somewhat as follows: In order to be strong and safe, or stronger and safer, the person wants to identify with and possess the symbols of a social group above that of his original family. In order to make this transition, however, certain prescribed routes must be followed. The person must have a talent which brings him in touch with and makes him useful to the group into which he wants entry. This talent could be intellectual, could be a facility for making money, could be beauty, could be an exceptionally loving and understanding personality. If an individual has the wish to change position but does not have such a talent or does not enjoy it to a sufficient degree, he may find it impossible to make the transition. He may find himself unable to establish the contacts which will enable him to learn the rituals of behaviour of the superordinate group. He may gradually come to know that, though 'the promised land' is in sight, he will never enter it. Meanwhile the group he is trying to leave punishes him for being 'different' and the group he tries to enter rejects him as presumptuous. The realisation, conscious or unconscious, that his campaign has failed may serve as a severe frustration and produce varying types of aggressive and compensatory behaviour. The resentment of the person who fails of mobility is likely to be severely punished and thus to create an acute anger-fear conflict.

Mobility conflicts which are unconscious

Except in one circumstance, which we shall come to in a moment, it does not seem likely that conflicts such as the one just described are engendered in early childhood. The conflict may nevertheless be unconscious. This unconsciousness of the elements of an adult conflict can arise because the mobile person gets little help from his society in labelling his behaviour. He is not told what he is trying to do, and he has no clear understanding of what the techniques are. If he hits on the means of mobility, it is, from his point of view, a matter of luck or accident. He is ordinarily not permitted to think that different social classes exist because the social beliefs which protect the class system forbid this recognition. Usually the mobile individual sees himself only as rising in some value such as 'wealth' or occupation, but he does not realise that his real mobility will be founded on a complex set of behavioural adaptations and changes in taste and outlook. Usually, therefore, the mobile person does not know what is happening to him while it is happening, does not know how he failed if he fails and does not know until 'afterwards' how he succeeded if he succeeds. This is a set of conditions which is bound to baffle and to arouse a confusion of angry, rebellious, apathetic and submissive responses.

Children of a mixed-class marriage

The one circumstance that we can see under which difference in social class can have an effect on a small child is the case where the child is born of a mixed-class marriage. If the mother is superordinate, she might in some ways 'look down' on the father, apologise for him and limit his usefulness as a model to her male child. Such a mother may be unduly 'ambitious' for her children, attempting to speed them over the landscape of childhood instead of allowing them to find their natural pace through it. She may get satisfaction in imposing early cleanliness training because it seems to her like a guaranteee of the future precocity of the child. She may inculcate the sex taboos strongly because she feels that the 'goodness' of the child in this respect will keep it out of 'bad company' and aid its development in the schools. She may handle its angry tendencies severely in the hope of making it amenable and yet urge it to highly competitive performance outside the home. One would predict that this kind of family training would

give a special colouring to the circumstances which ordinarily produce conflict in small children.

A child in a class-stable family with parents matched from the class standpoint would not ordinarily discover in the early years of life that there is any group 'above' its parents. During the formative period these parents would play their august roles, majestic in their competence and authority so far as the child could see. It would only be later in life, perhaps first during school days, that the child would learn that there are any people who look down upon it or its parents. Undoubtedly such knowledge would have some kind of effect on the career of the child, but we cannot say what the possible outcomes might be. We can be sure, however, that the evaluation put on the self and the family by the surrounding society will be a fact of importance in the developmental history of every child.

SUMMARY

Conflict itself is no novelty. Emotional conflicts are the constant accompaniment of life at every age and social level. Conflicts differ also in strength, some producing strong and some weak stimuli. Where conflicts are strong and unconscious, the individuals afflicted keep on making the same old mistakes and getting punished in the same old way. To the degree that the conflict can be made conscious, the ingenuity and inventiveness of higher mental life can aid in finding new ways out of the conflict situation. This applies to all emotional dilemmas, to those which survive from early childhood and to those which are created in the course of later life.

DEVELOPMENT IN THE INDIVIDUAL

C. S. Ford and F. A. Beach

THE various forms of sexual behaviour that have been [identified by scholars] do not appear in complete form in the newborn organism. Many of them, in fact, do not appear even in the reproductively mature individual unless the requisite amount of practice and experience has occurred. This chapter is primarily concerned with the ontogeny of sexual habits – with the development of sexual patterns in males and females of our own and other mammalian species. As far as human beings are concerned, the rate and direction of such development may be strongly affected by the social code of the group in which the individual grows up. However, regardless of cultural conditioning, the kinds of sexual responses that are possible are limited by physiological factors. Accordingly, we begin this chapter with a discussion of the physiological processes involved in sexual maturation.

PHYSICAL MATURATION

Directly beneath the brain in a special bony cavity lies the pituitary gland. The forward or anterior portion of this organ is often referred to as the 'master gland' because hormones which it secretes control the activity of several other glands located in different parts of the body. Certain of the anterior pituitary hormones that are known as 'gonadotrophins' stimulate growth and activity in the male and female gonads or sex glands. The pituitary secretes very little gonadotrophic material during infancy and childhood, but just prior to puberty the production and release of these hormones are increased. This is the initial physiological step towards the achievement of sexual maturity. The exact age at which the change occurs varies from one individual to the next, but it takes place on the average during the thirteenth or fourteenth year in boys and perhaps a year earlier in girls.

The effects of gonadotrophins upon the reproductive glands are twofold. The ovaries and testes respond to the crucial pituitary secretions by beginning to synthesise and secrete their own hormones. In addition, the sex glands start to produce mature germ cells – eggs in the case of the ovaries and sperm in the case of the testes. The increase in the secretion of sex hormones usually antedates by a year or more the appearance of mature sperm and eggs.

Evidence revealing the relationship between pituitary secretion and gonadal function comes from several sources. Analysis of the urine of pre- and post-adolescent boys and girls shows that increase in the concentration of pituitary gonadotrophins is closely correlated with the sex-gland changes. The same sort of measurements reveals an increase in the secretion of testicular and ovarian hor-

mones shortly after the rise in the level of gonadotrophic substances. Direct examination demonstrates that the reproductive glands become larger at this time, and microscopic study makes it plain that the cellular structure of these organs is altered simultaneously. The presence of motile sperm or of ripe eggs can be demonstrated directly (however, see subsequent sections of this chapter that deal with the first occurrence of ovulation).

Experimental and clinical investigations show that the adolescent changes in testes and ovaries must be preceded by activation of the pituitary function. If the pituitary gland is removed from experimental animals before puberty, the gonads do not develop and other signs of sexual maturation are absent. Similar symptoms accompany the failure of the normal pituitary function in human beings. However, animals deprived of this gland will develop normally if they are injected with pituitary extracts. Finally, it is possible to evoke precocious puberty in some species by pituitary administration. Male chicks only a few days old react to the injection of the appropriate pituitary hormones by growing a comb, crowing and attempting to mate like an adult rooster. The testes in these animals are strongly stimulated by the pituitary treatment and therefore secrete a sex hormone. It is the male hormone from the testes that causes the physical and behavioural changes characteristic of sexual maturity.

Precocious physical and behavioural maturity has been induced in rats, mice, dogs, monkeys and other mammals by supplying the immature organism with large quantities of gonadotrophic hormone from the anterior pituitary gland. Similar results can be produced in our own species. Some boys whose testes fail to descend into the scrotum have been treated with pituitary extracts. If too much gonadotrophic hormone is given, the testes may respond by secreting large amounts of the male hormone and this substance in turn induces premature development of the sexual organs and of the secondary sex characters which are defined below.

At the time of puberty several additional glandular changes occur. Although its function is not at all clear, it is known that the thymus gland grows smaller. The thyroid becomes increasingly active as puberty approaches and then grows somewhat less so in the later stages of adolescence. The behavioural consequences of these extragonadal endocrinological changes are of little interest at this point; hence they will not be considered further.

We have seen that the first step towards sexual maturation is increased pituitary function, and that the second is growth and maturation of the gonads, both as sources of a sex hormone and as the site of the production of fertile germ cells. The third phase in this chain of events consists of a variety of bodily responses to the sex hormones that are released from the reproductive glands. Some of these changes are internal and are largely unnoticed except by the experimentalist or the physician. Others are well known and recognised in every human society, and may in fact be defined as signals that call for changing the individual's social status.

The primary sex characters are the gonads – ovaries in the female and testes in the male. Accessory sex structures include the vagina, uterus and Fallopian tubes of the female, and the penis, seminal vesicles, prostate gland and epididymis of the male. One effect of the sex hormones is to produce growth of these accessory structures. In girls, the lips or labia of the vulva enlarge at the time of puberty and the uterus gradually assumes its adult form. The first menstruation (menarche) occurs in response to increased concentration of ovarian hormones, usually well in advance of the time that the entire reproductive tract is mature. In boys, the penis grows larger and the seminal vesicles and prostate increase in size and become functional. Not until the latter two organs are developed is the male capable of normal ejaculation, although, as noted below, orgasm can occur at a much earlier age.

The secondary sex characters are numerous and somewhat variable from species to species. In male and female human beings they include pubic and axillary (under-arm) hair. Deepening of the voice occurs in both sexes, although it is more marked in the male. Facial hair begins to appear in boys; other body regions including the chest, forearms and legs may also become more hirsute. The female breast is a secondary sex character and growth of this structure is usually rather rapid at the time of puberty. In males, the shoulders tend to become wider and in females the hips undergo similar alteration.

Proof that the sex hormones are responsible for these developmental changes in accessory and secondary sex structures is abundant. When, in human beings, the gonads fail to undergo the expected growth changes during puberty, the other bodily alterations described above do not occur. Men whose testes never

develop do not grow beards; they retain the high-pitched pre-adolescent voice and fail to acquire the normal masculine bodily conformation. Girls whose ovaries do not mature never menstruate and do not develop prominent breasts. If the reproductive glands are removed from domestic or experimental animals before puberty, the normal physical and behavioural signs of adolescence fail to materialise.

Many of the phenomena usually associated with puberty can be produced in human beings and lower animals by the administration of the appropriate sex hormones. Men without functional testes react to injections of a male hormone by developing a deeper voice, sprouting a light beard, and assuming more masculine bodily proportions. These changes have been produced in some individuals as late as the third decade of life. Women lacking normal ovaries can be caused to menstruate and to develop larger breasts if they are treated with estrin, one of the ovarian hormones.

Finally, many of the accessory and secondary sex characters can be caused to develop prematurely under the influence of sex hormones derived from some source other than the individual's own reproductive glands. Infant male animals treated with testicular hormone display excessive growth of the sexual organs and accentuation of the secondary sex characters, and in general take on the appearance of a miniature adult. Similar changes occur in immature females that are injected with ovarian hormones.

In human beings there sometimes occurs a condition known as precocious puberty. The outward signs of this abnormality are striking. Male children only two or three years of age may show all the physical signs of maturity. In some cases there is a heavy beard growth. The voice is deep and the accessory sexual structures are well developed. In normal human beings one part of the adrenal gland secretes hormones that are very closely related to the testicular hormones. Cases of masculine precocious puberty are usually caused by tumours of the adrenal which cause that gland to produce very large amounts of the male sex hormone.

The fact that the secondary and accessory sex structures in the male respond to the appropriate hormone injections, even though the testes may be immature, indicates that under normal conditions the changes which occur at puberty are due to the male hormone from the gonads. Somewhat similar conditions sometimes occur in female children. Little girls have been known to menstruate as early as three or four years of age, and in such instances other secondary sex characters tend to assume the adult condition. Most of these children suffer from tumours of the ovary which cause that gland to secrete ovarian hormones in amounts equal to those normally produced during adulthood. It is particularly instructive to note that despite the striking advance in physical development, children showing signs of precocious puberty are mentally and emotionally normal for their chronological ages. Furthermore, so far as is known, their sexual behaviour is childlike rather than adult. This fact is especially important in connection with the evaluation of the role of learning in normal human sexuality – a topic that is discussed later in this chapter.

DIFFERENTIATING PUBERTY FROM ADOLESCENCE

The terms 'puberty' and 'adolescence' have been used frequently in the foregoing discussion, and it is now necessary to consider their precise meanings. It is sometimes assumed that once an individual has passed through puberty he or she is sexually mature. Actually, however, puberty and maturity are temporally separated. Derived from the word *pubes*, meaning hair, the first term refers to the initial appearance of body hair in the genital region. It signifies the beginning of sexual maturation.

Adolescence is the period extending from puberty to the attainment of full reproductive maturity. Adolescence lasts several years in some species and involves a complex series of changes in several organs and organ systems. Different parts of the reproductive system reach their maximal efficiency at different stages in the life cycle; and, strictly speaking, adolescence is not completed until all the structures and processes necessary to fertilisation, conception, gestation and lactation have become mature. This complex maturational process may stretch over a period of ten years or longer.

The outward signs of puberty occur in most boys during approximately the fourteenth year. Rarely, however, do the testes contain fertile sperm any earlier than the fifteenth year and in many cases fertility may be delayed even longer. Menstruation starts in most girls before their ovaries are capable of producing ripe eggs. And egg production, in turn, begins before the uterus is

mature enough to support normal gestation. The attainment of full reproductive maturity is therefore a gradual process, subject, of course, to marked individual differences. For most individuals the menstrual rhythm during the first year or two after its inception is less regular than it will be in adulthood. Very often the period of flow is protracted and several months may intervene between the first few periods.

Relatively few girls are capable of reproduction before fifteen years of age, and even then their reproductive capacity is not as great as it will be later. The average age for full reproductive maturity in women has been estimated at approximately twenty-three years. This means that coitus is less likely to result in conception in the post-puberal girl than in the mature woman. A statistical study of age at first pregnancy in 2535 Scottish wives bears out this general conclusion. The data presented in Table 6.1

TABLE 6.1

PERCENTAGE OF WIVES CONCEIVING DURING THE FIRST TWO YEARS OF MARRIAGE

Number of women	*Age at marriage*	*Per cent conceiving*	
		First year	*First two years*
700	15–19	13·71	43·71
1835	20–24	18·49	90·51

show that women fifteen to nineteen years of age are less likely to become pregnant during the first two years of marriage than are wives who are five years older. (These figures are based upon the early findings of J. Mathews Duncan, as summarised by M. F. Ashley Montagu in 1946.)

The phenomenon involved here is customarily referred to as adolescent sterility. It may be the explanation for certain important anthropological observations having to do with sexual behaviour. In a number of human societies girls are permitted free sexual activity from puberty until marriage, but there appear to be relatively few conceptions even though contraception is reputedly not practised. The discrepancy between frequent copulation and infrequent impregnation is particularly striking in the case of the Ifugao natives of the Philippines and the Trobrianders. This matter is discussed in greater detail later in the present extract.

Some lower animals mature more rapidly and some more slowly than man. Laboratory rats and mice ordinarily become capable of fertile mating within about two months after birth. The age of reproductive maturity is approximately eights months in some breeds of dogs, three years in rhesus monkeys, nine years in female and seven and one-half years in male chimpanzees, and thirteen years in the elephant. A few of the larger mammals mature much less rapidly. For example, fertile mating does not take place in the rhinoceros until the twentieth year. All mammals that have been carefully studied pass through a developmental period of adolescent sterility similar to that which is believed to occur in human beings.

In rhesus monkeys, baboons and chimpanzees, as in the human female, the first menstrual cycles tend to be longer and less regular than they are in adult animals. The earliest menarche recorded for a chimpanzee is 7·5 years and the latest 10 years. The first few cycles in this species are usually 46 to 54 days long, whereas those of the healthy, mature female rarely exceed 36 days and usually average about 32 to 34 days. The period of adolescent sterility lasts from four months to more than two years in different individuals. Young and Yerkes collected data that show the relative fertility of adolescent and adult chimpanzees. The figures presented in Table 6.2 demonstrate that the differences seen in women have their parallel in the ape.

TABLE 6.2

RELATIVE FERTILITY OF ADOLESCENT AND ADULT FEMALE CHIMPANZEES

	Results of 78 pairings of younger females	*Results of 177 pairings of older females*
Per cent of matings resulting in pregnancy	14·0	30·0
Per cent of pregnant females aborting within 100 days or less after conception	36·0	18·0

PUBERTY CEREMONIES IN DIFFERENT SOCIETIES

In contrast to the gradual developmental changes that characterise the period of adolescence, the onset of puberty is comparatively

abrupt and has therefore claimed the attention of human beings in many societies. Special ceremonies are designed to mark this phase of the individual's life.

Girls

The onset of menstruation furnishes an obvious and dramatic signal of approaching physiological maturity in girls. In a large number of societies[1] the first menses are an occasion for more or less elaborate ceremonial. One common feature of many such ceremonies is the seclusion of the girl, particularly from any contact with men. The period of seclusion may last for only a few days or it may continue for several months.

In many instances the secluded girl receives special instruction from an older woman in matters pertaining to sex and marriage. The tuition generally includes an explanation of the social regulations governing proper conduct in sexual affairs, a description and sometimes a demonstration in pantomime of the techniques of love-making, advice on how to get along in married life, methods of avoiding conception and what to expect in childbirth. In some societies the pubescent girl must undergo certain hardships while she is in seclusion. She may be 'deflowered' and subjected to genital mutilation. Her ears may be pierced, her skin tattooed or scarified, her hair cut off and her teeth filed or blackened. Very often the conclusion of the period is marked by a feast or dance at which the girl, after bathing or going through ritual purification, publicly dons the clothes of a mature woman.

Some comprehension of the nature of puberty ceremonies for girls can be gained from the following examples. Among the northern Thonga of Africa, the girl at the time of her first menstruation goes to an older woman of her own choice and announces that she is of age. Subsequent events are described by Junod as follows:

> They will begin a *seclusion period* of one month. Three or four girls receive the initiation together. They are shut up in a hut, and when they come out, must always wear over their face a veil consisting of a very dirty and greasy cloth. Every morning they are led to the pool, and their whole body is immersed in the water as far as the neck. Other initiated girls or women accompany them singing obscene songs, and drive away with sticks any man who happens to be on the road, as no man is allowed to see a girl during this period. If a man happens to come near the group, the women ask him the secret formulae of the circumcision school, not the long ones but the short ones, probably those which contain licentious words. Should he be unable to answer, they beat him. It is said that a man who sees a girl during this month becomes blind! When the cortège of women accompanying the initiated has returned home, the nubile girls are imprisoned in the hut. They are teased, pinched, scratched by the adoptive mothers or by other women; they must also listen to the licentious songs which are sung to them. Though they are trembling from cold, being still wet, they are not allowed to come near the fire. They are also instructed in sexual matters, and told that they must never reveal anything about the blood of the menses to a man. They are also exhorted to be very polite to every grown-up person, and must salute everybody entering the hut, even those passing before the door, by clapping their hands. Sometimes the wind moves some dead leaves; they mistake this noise for the sound of steps and salute reverently!
>
> At the end of the month the adoptive mother brings the girl home to her true mother. She also presents her with a pot of beer. A feast takes place on this occasion. . . .

At menarche the Marshallese girl retires to a specially built hut, accompanied by a number of immature girls who are to act as her maids-in-waiting. The pubescent girl wears two loin mats in the fashion of an adult woman and underneath them a wadding to absorb the menstrual blood. For fifteen days she remains secluded and during this time she may not eat fish or meat. All cooking for the girl and her attendants must take place outside the hut. They are looked after by an elderly woman who teaches them songs and tells them moralistic stories. Every morning before dawn the girls go down to the sea to bathe and put on clean mats. When they return to the hut the elderly woman purifies and perfumes the pubescent girl by burning sweet-smelling leaves and passing them between her legs. On the fifteenth day a final ceremony takes place which releases the girl from seclusion and marks the end of her girlhood.

Lenge girls go through an elaborate puberty ceremonial. They

are 'deflowered' with a special tool made of horn and remain in an initiation school for an entire month during which they may not eat meat or salt. Sexual instruction is given indirectly through the recitation of magical formulas. At the close of the period a day is spent bathing in the sea. Fertility medicine is given to the girls and they then return home to remain in seclusion for two or more months.

Practices similar to those occurring at the time of puberty take place in some societies at times other than at menarche. Attempts to enlarge the labia, ear piercing, tattooing and scarification may take place either before or sometimes after first menstruation. Among the Alorese, for example, girls are tattooed at any time between the ages of ten to fourteen. Before puberty, girls on Ponape undergo treatment designed to lengthen the labia minora and to enlarge the clitoris. Old impotent men pull, beat and suck the labia to lengthen them. Black ants are put in the vulva; their stinging causes the labia and clitoris to swell. These procedures are repeated until the desired results are attained. At the age of about fifteen Jukun girls have their ears pierced during a minor ceremony. Prior to this time and for three weeks thereafter they may not indulge in sexual intercourse. When Venda girls attain the age of ten to twelve years they begin to practise the custom of pulling out and lengthening the labia minora. Girls who fail to do so are criticised as being lazy. 'You will remain nothing better than a post one smears one's dirt on. You are a tree that lends no hold. Just a hole without anything.'

Like 'coming-out parties' in our society, puberty ceremonies for girls serve among other things to announce to the public that these young women have now attained adult status. In some societies the initiation rite marks the point at which girls are permitted to begin indulging in sexual intercourse. If sexual life for girls begins before puberty, such ceremonies may signalise the young woman's eligibility for marriage.

Boys

A number of societies provide initiation rites for boys. In some cultures the only ritual is the adoption of some symbol of manhood. On Truk, for example, when the boy's pubic and facial hair becomes sufficiently noticeable, he merely adopts a red loin cloth and goes to live in the men's dormitory. Among some other peoples, however, the ceremonial is very complicated and extends over a considerable period of time.[2] For boys there is no event comparable to menarche in girls that can be seized upon as an indication of maturity. Nevertheless, such physical signs as the appearance of facial and other bodily hair or the first nocturnal emission are sometimes taken as an indication of masculine maturity, and boys may at that time undergo an initiation ceremony. On the other hand, many of the societies which subject boys to complex initiation ceremonies pay little attention to the time at which the rite takes place. Quite frequently boys are initiated in groups, and the ceremonies are conducted at any time when enough boys are considered sufficiently mature.

In a large number of societies, initiation for boys consists primarily of a period of seclusion during which the neophyte undergoes ordeals and receives instruction on sexual matters. In some, genital mutilations including circumcision, subincision (slitting a section of the urethra), and in one society at least (Ponapeans) hemicastration (the removal of one testis), form a part of the initiation ceremony. Sodomy constitutes part of the puberty initiation for boys in a few societies. Women are usually excluded from these ceremonies, and the practices which take place are carefully kept secret from all feminine members of the society. The following examples illustrate the nature of initiation ceremonies for boys.

Among the Keraki, boys undergo a most complicated initiation, which we can only summarise briefly. The initiates are several in number, about thirteen years old. In preparation for the period of initiation a space is cleared in the bush, and by the edge of the clearing a structure is built as a dormitory. The first important rite is the revelation of the bull-roarer (a primitive musical instrument) to the uninitiated; while this is being done each boy must remain perfectly silent and motionless. Following this the boys submit to a blow on the back with a heavy banana stalk. Next come a feast and a parade through the village, after which the initiates return to the dormitory. That night begins a period of nearly a year during which the boys play a passive role in sodomy with older men. During this time they may not even be seen by women but are constantly at the service of older men, who may be fellow villagers or visitors wishing to have anal intercourse with the initiates. At the end of the period a second feast occurs and the

boys are released from seclusion to return home, entitled now to assume the role of an adult male.

When the Chewa youth reaches the age of puberty the chief of his village is advised of this fact by the boy's maternal uncle. Preparations are made for the initiation, which is usually delayed until several boys are ready to undergo the ceremony. When the day arrives the boys are led blindfolded to a secluded spot where the cult figures are made and kept. Of the boy's treatment there Steytler writes:

> Here the boy is kept for about a week and he is cruelly treated. He is swung over a slow burning fire, drawn prostrate over the hard ground, beaten with rods, rolled in ants and itching beans, pulled about with string tied to his private parts, made to eat excrement and drink urine, etc. The test is very severe and if the boy in pre-puberty days happened to be disobedient . . . he is more severely treated still. . . . When everything is over and past the boy is taken to the chief, presents are exchanged between relations and the chief, and the boy receives his new name.

DEVELOPMENT OF SEXUAL BEHAVIOUR IN HUMAN BEINGS

It has been explained that the attainment of sexual maturity is a gradual process and that the various organs and organ systems involved in reproduction become functional at different stages in the life of the individual. We have further noted that in many societies the individual's entrance into adolescence receives public recognition in the form of puberty ceremonials. But up to this point very little has been said about sexual behaviour *per se*. The nervous and muscular mechanisms involved in sexual arousal and its overt expression can properly be classified as very important elements in the interrelated system of physiological factors that must become mature before reproduction can occur.

Reflexive components

Some of these mechanisms for behaviour are reflexive, whereas others develop only as a result of practice and learning. For example, the human male does not have to learn how to fill his penis with blood so that it becomes erect and rigid, but he may have to learn how to copulate; the conditions under which an adult man experiences erection undoubtedly are influenced by his life experiences. There are decided differences between various animal species as regards the extent to which mating responses depend upon or are modified by learning and conditioning.

It is possible to analyse separately the several reflexes and more complex reactions that normally appear simultaneously or sequentially in the sexual act of adult males and females. When this is done it becomes obvious that different segments of the total response mature at different rates. In the human male, for example, complete genital erection is possible from the day of birth, and baby boys frequently show this reaction under the influence of bladder distension or in response to manipulation of the phallus. The power of ejaculation, in contrast, is not acquired until puberty, and the production of normal sperm is delayed until even later in life.

In grown men orgasm and ejaculation usually occur together, but they are not necessarily mutually dependent. As a matter of fact, it is reported that sexual climax or orgasm can be produced in very young human infants of either sex. Kinsey, Pomeroy and Martin state that male infants less than one year of age respond to manipulation of the genitals by making thrusting movements with the pelvic muscles; and if the stimulation is continued the baby's movements become more rapid and vigorous and culminate in a general spasm quite similar to that which characterises climax in most adults.

The evidence therefore suggests that the neuromuscular system of the human animal is capable at birth of mediating at least two of the basic reflexive patterns that will later be woven into the complete sexual act. But adult patterns of sexual behaviour consist of a great deal more than genital and pelvic reflexes. What other mechanisms are involved, how do they develop, and when do they become mature? How are the simple, inherited reflexes elaborated into the much more complex and variable forms of sexual expression that characterise adults in different societies? A fully satisfactory answer to these questions could be obtained only by studying the sexual play of many children of varying ages in a large number of different cultural settings. Unfortunately, however, the amount of such activity that occurs and the ease with

which the observer can obtain information concerning it are controlled to a large extent by the attitudes of adults. And these attitudes vary to a great extent from one society to the next.

Restrictive societies

In a minority of the societies concerning which we have adequate information adults attempt to deny young children any form of sexual expression.[3] As will be explained later, this is the prevailing attitude in American society, although there is considerable variance between actual behaviour and the idealised standards of the moral code.

The severity of restrictions and punishments associated with sexual transgressions in childhood varies from one restrictive society to another. Among the Apinaye, for example, boys and girls are warned from infancy not to masturbate and a severe thrashing awaits the child suspected of such behaviour. In Africa, Ashanti boys are told by their fathers at an early age not to masturbate or engage in any sexual play. In New Guinea, Kwoma boys are constantly warned not to finger their genitals; if a woman sees a boy with an erection she will beat his penis with a stick, and boys soon learn to refrain from touching their genitals even while urinating. Kwoma girls also are told not to finger their genitals but are not punished for so doing. The Cuna specifically forbid their children to engage in either homosexual or heterosexual play; and youngsters among the Chiricahua are whipped if they are detected playing sex games.

Most of these restrictive societies maintain a public conspiracy against the acquisition of any sexual knowledge by children. Adults avoid mentioning matters of sexual significance in their presence, and make every attempt to keep them in total ignorance of the reproductive process. Among the natives of the western Carolines sex is never discussed before children, especially girls. Cuna children remain ignorant of sexual matters (as far as adult instruction is concerned) until the last stages of the marriage ceremony. They are not even allowed to watch animals give birth. Chagga children are told that babies come out of the forest.

In a number of these societies particular pains are taken to prevent young children from accidentally observing sexual behaviour. In some instances, as among the Murngin of Australia, boys are removed from the dwelling to the boys' house or bachelors' hut when they are four or five years old; this is done for the specific purpose of preventing them from witnessing sexual behaviour at home. The Kwoma husband and wife are always careful to wait until the children are asleep before indulging in sexual intercourse.

Such adult attitudes towards childhood sexuality may prevent youngsters from engaging in sexual practices in the presence of their elders, but whether they successfully suppress sexual activity in secret is another matter. There is evidence that in some of these societies children do engage in a certain amount of sexual behaviour despite strong adult disapproval. In Haiti little boys and girls privately experiment in sexual activity from early childhood until puberty. Manus children masturbate, but always in solitude and surrounded by shame. When they are alone in the bush Kwoma boys scrape the penis with nettles. And on Truk, children play at intercourse at an early age, although their parents will beat them if they are caught. In Trukese society children do sometimes observe their elders engaging in sexual activities at night. Apinaye boys and girls masturbate frequently even though such play is punished whenever it happens to be observed, and despite the fact that at a ceremony which is conducted when they are half grown their genitalia are examined and the children are flogged if there appears to be evidence of masturbation. In the case of boys, this 'evidence' is described as 'retractibility of the prepuce'. But the validity of such criteria is questionable. Actually, there are no known physical stigmata that constitute reliable evidence of habitual masturbation as far as the male is concerned. R. L. Dickinson has long held that prolonged habits of feminine masturbation involving vulvar traction and friction leave permanent signs in the form of lengthened and corrugated labia. However, Dickinson's thesis is not accepted by all authorities, and in any event it applies to mature women and not to 'half-grown' girls.

Some peoples make a sharp distinction between socially immature and mature persons with respect to permissible sexual activity. These societies take the attitude that sexual intercourse before adulthood must be avoided; but once the person is mature by their standards, considerable freedom in sexual matters may be allowed.[4] For the most part these peoples seem particularly concerned with the pre-pubescent girl, believing that intercourse

before the menarche may be injurious to her. Girls of the east central Carolines are strictly forbidden intercourse before puberty, but after that they enjoy almost complete sexual freedom. After menarche, Ao girls begin to sleep in dormitories where they indulge in intercourse with partners of their choice. Among the Siriono intercourse before puberty is forbidden, but premarital affairs are customary once the girl has menstruated. The Chukchee believe that intercourse will harm a girl until her breasts are fully developed or until she begins to menstruate. However, immature girls often engage in coitus despite this belief. In this society it is considered proper for a girl to carry on serious love affairs between the time of the first menstruation and marriage. The Ashanti are convinced that sexual intercourse with a girl who has not undergone the puberty ceremony is so harmful to the community that the offence is punishable by death for both partners. Premarital intercourse is also forbidden to the postpubescent Ashanti girl, but this rule is not nearly so strictly enforced.

In most of the African societies in our sample[5] boys are strictly forbidden to have intercourse before undergoing the puberty ceremony or initiation rite. The Chagga boy, for example, cannot have intercourse until he has been circumcised and properly initiated into adult status. If caught in the act the boy and his partner are laid one on the other and staked to the ground. After circumcision all Chagga boys have intercourse with a barren woman and subsequently they sally forth to seek other sex partners. Until marriage they are instructed to practise either interfemoral intercourse or coitus interruptus unless the girl places a pad in the vagina to avoid conception. A comparable attitude is taken by the Jivaro of Ecuador, who strictly forbid boys to engage in intimate relationships with girls until they have gone through an initiation ceremony at puberty.

In other societies the prohibitions against sexual intercourse continue unabated or, in some instances, are intensified after puberty and remain in force until marriage or at least until betrothal. The methods used to prevent premarital sexual activity during adolescence include segregation of the sexes, strict chaperonage of girls and threats of severe disgrace or physical punishment. The extreme pains to which adults in these societies are forced to go in order to control the sexual behaviour of young people is an eloquent expression of the strength of the tendency on the part of older children and adolescents to engage in such activity. There are indeed very few societies in which any method of control appears to be completely effective in preventing heterosexual intercourse among young unmarried couples.

Perhaps the most nearly successful method of controlling the sexual activity of young people is to separate the sexes and keep the girls under constant surveillance. Among the Abipone, for example, boys and girls were strictly segregated at all times and premarital chastity is said to have been universal. A similar situation exists among the Arapaho, Cheyenne, Papago and Wapisiana, all of whom keep the sexes strictly apart from childhood. Boys and girls never play together, and until marriage young men and women never associate in the absence of chaperones. The only completely effective prevention of premarital relations has been devised by the Wapisiana. They define cohabitation as marriage, and thus rule out the possibility of intercourse between two unmarried people. At the same time, of course, they eliminate the problem of the unmarried mother.

In most of the societies that practise segregation and chaperonage to control the sexual behaviour of adolescents, boys are less carefully watched than girls; and, in some cases at least, it appears that youths are able to circumvent the barriers, with the result that sexual intercourse before marriage not infrequently occurs. For example, among the Hopi a strong attempt is made to keep boys and girls apart from the age of ten until marriage; the girls are kept at home and are accompanied by an older woman whenever they go out. Girls are expected to be chaste until marriage. Boys, however, are not similarly restricted, and, whenever possible, they defeat the chaperonage system by crawling into the girl's house stealthily at night or by holding clandestine prearranged meetings. The Hopi place all the blame for illegitimate pregnancy squarely upon the girl who is involved. Her friends ignore her and her family scolds her, but the lover is not regarded as at fault and is neither forced nor expected to marry her.

A similar situation exists among the Kiwai Papuans of New Guinea. There, girls are carefully chaperoned by their parents and usually kept in ignorance of love-making for some time. The boy, however, is not similarly restricted and will take the initiative in attempting to get around the rules. The young pair have to be

very clever to meet; they usually are able to do so only at night. The girl may slip out of the house after her parents are asleep or the boy may sneak into her house through the floor. And apparently, despite every effort on the part of adults, many couples find it possible to carry on love affairs in secret.

Threats of the most severe disgrace and punishment do not appear to be completely effective in preventing young people from engaging in sexual activity before marriage. In the Gilberts, for example, great emphasis is placed on a girl's chastity before marriage. If a girl is seduced and it becomes public knowledge, both parties are put to death. Nevertheless, the evidence indicates that many transgressions take place in secret. In only a few societies – namely, the Vedda, Keraki, Chiricahua and Sanpoil – is the burden of guilt placed upon the boy. Among the Vedda, for example, if a man were seen even talking to an unmarried girl her relatives would kill him. In most of these restrictive societies, however, the threats of disgrace and punishment are specifically directed towards misbehaviour on the part of the girl.

The attempts of adults to restrict the sexual behaviour of adolescents seem, in many of these societies, to be intended primarily as a means of insuring the virginity of the unmarried girl. Some peoples attempt to determine whether or not a girl has remained chaste by conducting a crude examination of her sexual organs. The hymen or 'maidenhead' is a tab of tissue which, in most virgins, partially obstructs the entrance of the vagina. Very often this structure is nicked or stretched when the vagina is sufficiently penetrated for the first time, and a certain amount of blood may be lost. However, the size and thickness of the hymen vary from individual to individual. As a result, some fully virginal girls may bleed very little if at all during the first intercourse, whereas in other women a great deal of stretching is necessary before the obstruction offered by this tissue is completely removed.

Despite the actual unreliability of their tests, some societies use the occurrence of bleeding in response to vaginal penetration as an important index to chastity. Among the Kurd, for example, when the bridegroom has intercourse with the bride the nuptial cloth is examined for blood. If a girl is shown to have been a virgin, the cloth is paraded on a stick through the village and the bride price is then paid. If, however, the bridegroom finds himself disappointed, the girl is heaped with abuse, given back to her parents and in some instances subjected to further public disgrace. Yungar girls are 'deflowered' a week before marriage by two old women. If, at this time, examination fails to indicate chastity, severe penalties are meted out to the girl, including starvation, mutilation, torture and even death.

American society

The social code pertaining to sexual behaviour of children and unmarried adolescents in the United States is clearly a restrictive one. In this country, constant pressure is exerted, ideally at least, to prevent any form of sexual behaviour until it is legalised and can occur in the bonds of matrimony. In this society, as in many others, there is a tendency towards a double standard in respect to premarital sexual behaviour. More pressure is brought to bear upon unmarried girls than upon boys. In actual practice, furthermore, the burden of protecting young persons from indulging in sexual activity falls somewhat more heavily upon the parents of American girls than upon the parents of boys. This tendency towards a double sex standard is also reflected in attitudes towards extramarital sexual activity. As has been shown, a double sex standard during late childhood and adolescence is characteristic of many societies, but not of human beings in general. There are societies in which there is little if any difference in the premarital sexual restrictions placed upon girls and boys, and in a few societies the more severe restrictions confine the male rather than the female.

Kinsey and his associates have shown in statistical fashion what has generally been recognised for many years, namely, that in spite of the attitudes of adults, children in our society frequently indulge in many forms of sexual activity. Although the strictness with which the moral code is enforced varies considerably from one social class to another, a more or less concerted attempt to prevent children from indulging in any form of sex play continues well into adolescence and up to the time of marriage. And, as we have noted, most other societies that discourage infantile and childhood sex play also attempt to control premarital experimentation in sexual matters on the part of adolescents or young adults.

Although this attitude is characteristic in America, the strength of condemnation varies somewhat from one social group to another. However, regardless of the prohibitions against it,

premarital sexual behaviour does occur in a fairly large proportion of the population. It does not necessarily involve actual copulation. Landis and his co-workers report that 59 per cent of the married women whom they questioned had indulged in extensive heterosexual play without coitus prior to marriage. Forty-two per cent of the unmarried women in this study admitted the occurrence of sex play.

According to Kinsey, Pomeroy and Martin, more than 80 per cent of American males engage in petting before they are twenty years old. By the age of twenty-five, approximately one-third of the male population has achieved orgasm in this fashion. Men from the higher educational levels are likely to confine their adolescent activity to non-coital techniques, whereas individuals from the lower socio-educational strata tend to proceed more or less directly to coitus and indulge in a minimum of petting.

As far as actual copulation is concerned, Kinsey and his collaborators report that it was attempted during or before adolescence by 22 per cent of the American boys they interviewed. The first experiment usually takes place between the ages of 10 and 14. By the time he is 12 years old approximately one boy in every four or five has at least tried to copulate with a girl or woman. More than 10 per cent of these youths experience their first ejaculation in connection with heterosexual intercourse. Considering all the men interviewed in Kinsey's study, it becomes apparent that more than two-thirds of them had at least one premarital experience involving copulation. The incidence of such behaviour varies with the individual's socio-educational level, being least frequent in college-educated groups, and nearly universal among men who have no more than an eighth-grade education.

Terman found that approximately one half of 760 American husbands whom he studied admitted premarital intercourse with the women they later married. Seven per cent of this group said they had copulated with at least one other woman prior to marriage, and 26 per cent mentioned intercourse with five or more women before marriage. Only 13·3 per cent of the 777 wives represented in Terman's sample admitted premarital relations with the husband, and much smaller percentages listed intercourse with other males before marriage. The lack of agreement between accounts given by husbands and wives probably reflects chiefly a different degree of resistance to confessing premarital freedom. Terman was particularly impressed with differences between older and younger married couples as regards premarital behaviour. He noted that the proportion of men and women who were virgins at marriage had steadily decreased between the approximate dates of 1910 and the early 1930s:

> *If the drop should continue at the average rate shown for those born since 1890*, virginity at marriage will be close to the vanishing point for males born after 1930 and for females born after 1940. It is more likely that the rate of change will become somewhat retarded as the zero point is approached and that an occasional virgin will come to the marriage bed for a few decades beyond the dates indicated by the curves. It will be of no small interest to see how long the cultural ideal of virgin marriage will survive as a moral code after its observance has passed into history.

Semi-restrictive societies

There is no clear-cut dividing line between restrictive and semi-restrictive societies; and often, as we have pointed out, the sexual codes that adults attempt to enforce on immature members of the group differ according to the young person's sex or age. There are, however, many societies[6] in which the adult attitudes towards sex play in children or towards premarital affairs in adolescents are characterised by formal prohibitions that are apparently not very serious and in fact are not enforced. In such cases sexual experimentation may take place in secrecy without incurring punishment, even though the parents know perfectly well what is going on. The Alorese formally object to any form of sex play on the part of older children. But overt homosexual and heterosexual practices on the part of boys and girls occur, and children playing together in field houses imitate the sexual intercourse of their parents. Unless this is brought flagrantly to the attention of the adults they do nothing about it.

Among the Andamanese premarital promiscuity is common and the parents do not object as long as the love affairs are kept secret. Parents object to such activities in theory, but unless they are practised openly no punishment is involved. Should a girl become pregnant, however, the parents of the couple usually arrange for them to be married. The Huichol uphold an ideal of premarital

chastity for both sexes, but in practice this is rarely realised. If an adolescent couple is caught in sexual intimacy, both individuals are beaten and they are forced to marry; but parents do not keep close surveillance over the activities of young people, and the latter have many opportunities to slip off into the bush in the evening during feasts and dances.

In some societies the only recognised sign of sexual transgression on the part of young people is premarital pregnancy. Among such peoples it appears that intercourse frequently takes place between unmarried couples, but numerous devices and techniques are employed either to prevent conception or to abort an unwanted foetus. In a number of African societies it is customary for adolescent boys to practise interfemoral intercourse or coitus interruptus to avoid impregnating the girl. Contraceptive measures used by young people in these societies include placing a pad of absorbent material in the vagina, washing the passage after intercourse, and orally ingesting certain medicines believed to ensure temporary sterility. Should these fail to prevent conception, the girl may resort to an abortion.

Permissive societies

Adults in a large number of societies take a completely tolerant and permissive attitude towards sex expression in childhood.[7] Under such conditions youngsters engage in a certain amount of sexual play in public. The fingering of the child's own genitals follows exploratory movements of the hands which contact the various parts of the body. If adults do not attempt to discourage such behaviour, fingering the genitals becomes an established habit of occasional occurrence. As the child grows old enough to walk about and play with others, he tends to extend the range and to increase the variety of sexual activities. Handling the genitals of others of the same or opposite sex occurs frequently under conditions of free sex play. Additional forms of sexual activity on the part of young children sometimes include oral-genital contacts and attempted copulation with a sex partner.

In a few permissive societies adults participate actively in the sexual stimulation of infants and young children. Hopi and Siriono parents masturbate their youngsters frequently. And in these societies self-masturbation passes practically unnoticed during early childhood, adults taking a tolerant and permissive attitude towards all sexual behaviour at least until the age of puberty. Among the Kazak, adults who are playing with small children, especially boys, excite the young one's genitals by rubbing and playing with them. In this society autogenital stimulation on the part of young children is accepted as a normal practice. Mothers in Alorese society occasionally fondle the genitals of their infant while nursing it. During early childhood Alorese boys masturbate freely and occasionally they imitate intercourse with a little girl. As the children grow older, however, sexual activity is frowned upon and during late childhood such behaviour is forbidden in both boy and girl. Actually, however, they continue their sexual behaviour, but in secret.

Among the Pukapukans of Polynesia where parents simply ignore the sexual activities of young children, boys and girls masturbate freely and openly in public. Among the Nama Hottentot no secret is made of autogenital stimulation in early childhood. Young Trobriand children engage in a variety of sexual activities. In the absence of adult control, typical forms of amusement for Trobriand girls and boys include manual and oral stimulations of the genitals and simulated coitus. Young Seniang children publicly simulate adult copulation without being reproved; older boys masturbate freely and play sexual games with little girls, but the boys are warned not to copulate on the grounds that this behaviour would weaken them. Lesu children playing on the beach give imitations of adult sexual intercourse, and adults in this society regard this to be a natural and normal game. On Tikopia small boys induce erections in themselves through manual manipulation, and this is ignored or at most mildly reproved by adults. Little girls also may masturbate in this society without being punished for such behaviour.

Most of the societies that permit children free sex play (and some that are semi-restrictive) also allow them opportunity to observe adult sexual behaviour and to participate in discussions of sexual matters.[8] Among the Alorese sex knowledge is completely accessible to young children and by the age of five they are well informed on all details of the entire reproductive act. All members of the Pukapukan household sleep in the same room under one mosquito net; and although some parents wait until they think the children are asleep, there are frequent opportunities for youngsters to observe adult sexual activities and sexual matters are often

talked about. Lesu children are free to observe adults copulate, with the specific exception that they may not watch their own mothers having intercourse. On Ponape children are given careful instruction in sexual intercourse from the fourth or fifth year. Trukese children receive no formal tutelage, but they learn a great deal by watching adults at night and by asking their elders about sexual matters. Among the Wogeo sexual matters are freely discussed by adults in the presence of children. In this society, however, parents take some precautions against their own children observing them in intercourse.

In the societies where they are permitted to do so, children gradually increase their sexual activities both as they approach puberty and during adolescence. There are, indeed, some societies in which enforcement of the prevailing incest regulations is the only major restriction on sexual activity among adolescents.[9] As in the case of very young children, their sex play first includes autogenital stimulation and mutual masturbation with the same and opposite sex, but with increasing age it is characterised more and more by attempts at heterosexual copulation. By the time of puberty in most of these societies expressions of sexuality on the part of older children consist predominantly of the accepted adult form of heterosexual intercourse, the pattern which they will continue to follow throughout their sexually active years of life.

Among the Chewa of Africa parents believe that unless children begin to exercise themselves sexually early in life they will never beget offspring. Older children build little huts some distance from the village, and there, with the complete approval of their parents, boys and girls play at being husband and wife. Such trial matings may extend well into adolescence, with periodic exchanges of partners until marriage occurs. The Ifugao head-hunters of the Philippines maintain a similar attitude towards the sex play of older children and adolescents. In this society unmarried individuals live in separate dormitories from early childhood. It is customary for each boy to sleep with a girl every night. The only check on promiscuity is that imposed by the girls themselves. Usually a girl is unwilling to form too prolonged an attachment to one boy until she is ready to be married. Boys are urged by their fathers to begin sexual activities early, and a man may shame his son if the latter is backward in this respect. Even after puberty there seem to be relatively few instances of conception resulting from this free sexual activity. Pregnancies do occasionally occur, however, and in that event one of the girl's lovers must marry her.

The Lepcha of India believe that girls will not mature without benefit of sexual intercourse. Early sex play among boys and girls characteristically involves many forms of mutual masturbation and usually ends in attempted copulation. By the time they are eleven or twelve years old, most girls regularly engage in full intercourse. Older men occasionally copulate with girls as young as eight years of age. Instead of being regarded as a criminal offence, such behaviour is considered amusing by the Lepcha. Sexual life begins in earnest among the Trobrianders at six to eight years for girls, ten to twelve for boys. Both sexes receive explicit instruction from older companions whom they imitate in sex activities. Sex play includes masturbation, oral stimulation of the genitals of the same and opposite sex, and heterosexual copulation. At any time a couple may retire to the bush, the bachelor's hut, an isolated yam house or any other convenient place and there engage in prolonged sexual play with full approval of their parents. No marriage is consummated in Trobriand society without a protracted preliminary period of sexual intimacy during which both sincerity of affection and sexual compatibility are tested. Premarital pregnancy is said to be rare in this society, despite postpubertal sexual intercourse over a period of three years or more before marriage. This experience has led the Trobrianders to doubt a causal relationship between coitus and conception. Instead they consider supernatural influences to be far more significant in causing a child to be conceived.

In this instance, as in other cases of frequent but infertile coitus among postpubescent males and females, the phenomenon of adolescent sterility would appear to be particularly pertinent. It may well be that although they have passed the menarche, the girls involved in this activity are not yet ovulating, or at least are incapable of carrying a foetus to term. Any such interpretation must remain speculative, however, until there is more satisfactory proof for the absence of any form of contraception.

An interesting attitude towards the sexual activity of adolescents is taken by the Ila-speaking peoples of Africa. Childhood is regarded as a time of preparation for adult life and mature sexual functions. At harvest time each girl is given a house to which she takes a boy of her choice, and there they play as man and wife. It

it reported that there are no virgins among these people after the age of ten. On Easter Island children from the age of six on imitate the sexual behaviour of adults without censure; and young people among the Maori play together at being husband and wife at night in the bush. Full copulation frequently occurs before puberty. Lesu adults regard as natural the attempts at intercourse in which children engage, and they give full approval to free sexual activity on the part of adolescents.

DEVELOPMENT OF SEXUAL BEHAVIOUR IN SUB-HUMAN SPECIES

Unhampered by social restrictions, young animals of many species display sexual behaviour similar to that seen in human children in permissive societies. This is particularly true in the case of the other primates, the anthropoid apes and monkeys.

Sub-human primates

Systematic observations of infant and adolescent chimpanzees have shown that they engage in a great deal of sexual play. Although reproductive maturity is not attained until the eighth year or later, three-year-old chimpanzees show many autoerotic, homosexual and heterosexual responses. Masturbatory play appears in both sexes prior to puberty. According to Yerkes, solitary masturbation is more common in the male; when females show comparable behaviour they tend to do so in association with companions.

Like human children, little male and female chimpanzees play sexual games involving erotic advances by either or both partners and including attempts to effect actual copulation. Several different copulatory positions are experimented with and partial intromission is sometimes achieved. Some young males practise manual and oral stimulation of the feminine genitalia and females may handle the erect penis of their immature partners.

Other primates also engage in a great deal of sex play during infancy and childhood; in fact Zuckerman concludes that before puberty the young monkey reproduces all the sexual activities of its elders, as far as is physically possible. This worker studied the development of sexual patterns in several primate species. He describes a six-months-old male pigtailed monkey that formed a habit of mounting his mother in copulatory fashion whenever she adopted the coital posture. When seven months old this animal displayed erections and pelvic thrusts while covering the adult female. At this time the young one was still nursing; his age was roughly equivalent to that of a two-year-old child.

All manifestations of sexual behaviour are present in male and female baboons by the ninth month, at which time the milk teeth have just appeared. Carpenter reports observations of autogenital manipulation in infant spider monkeys living in their native environment. During the first year of life the captive male macaque displays sexual reactions to other animals and even to inanimate playthings. In the course of infantile play female monkeys often execute the adult sexual presentation. Immature animals of either sex may react sexually to the advances of a full-grown partner. Pre-adolescent female baboons may form lasting sexual partnerships with an adult male long before reproduction is a physical possibility.

Lower mammals

Males of many sub-primate mammalian species frequently execute portions of the adult copulatory pattern long before they are mature. In the course of the mock fighting, wrestling and chasing which characterise the pre-adolescent animal, young males often grasp another individual and make one or two coital movements before dismounting. This type of behaviour has been observed under laboratory conditions in rats, guinea pigs and golden hamsters, and comparable reactions are seen in many domesticated species. The male lamb may mount other young sheep as early as the first week of life, although puberty is not attained until several months later. Male calves react similarly, and so do male lion cubs. In no instance is the behaviour complete, for the ejaculatory response is lacking, and these fragmentary mating reactions do not give the impression of intense arousal or satisfaction. Instead, they appear to be parts of a broader pattern of play involving generalised excitement.

It is most significant that female mammals of sub-primate species do not display any of the adult feminine sex pattern prior to puberty. Only with the occurrence of the first heat period do the female's sexually receptive responses appear. This is in marked contrast to the frequent occurrence of masculine reactions in males

of the same species and the fact that immature female primates are as active sexually as males. The cause of these differences is not fully understood, but some clues are available.

FUNCTIONAL SIGNIFICANCE OF EARLY SEX PLAY

Having noted that sex play is common among young animals of many species and among human children as well, it is reasonable for us to inquire concerning the possible significance of this phenomenon. For example, does early practice prepare the individual for successful sexual performance in adulthood? Does each animal have to learn how to court and mate?

The answer depends upon the species and sex of the individual. As far as females of species below the primates are concerned, the answer is obvious. They do not practise the adult feminine mating pattern before puberty; nevertheless, when the first oestrus occurs they are capable of effective coitus. In several experiments males of the same species have been reared in isolation from other animals from the time of weaning until the attainment of maturity. Despite the lack of opportunity to engage in prepubertal sex play, these animals can copulate effectively the first time they are placed with a receptive female. Apparently no practice is necessary for successful coitus.

The copulatory behaviour of male primates is more dependent upon learning than is that of lower animals. Adult monkeys and chimpanzees that have had no heterosexual experience often are unable to copulate with the oestrous female. The inexperienced male may be sexually aroused and may attempt to mate when the opportunity arises; but even when the female is thoroughly receptive, the male seems incapable of fulfilling his sexual role. Often he starts to cover the female and then is so awkward in attempting intromission that the mating is never completed. As far as the male chimpanzee is concerned, several months or even years of practice and experience in sexual performance appear essential to the development of maximal coital efficiency.

It is somewhat surprising that similar behavioural deficiencies have never been observed in the mating of sexually inexperienced female monkeys and apes. When they are fully receptive, naïve females invite, accept and respond appropriately to the sexual advances of an experienced male. And if equally inexperienced males and females are put together after puberty, the female is obviously better prepared to carry out her part of the coital relationship. As a matter of fact she not only performs all the necessary feminine responses, but may attempt to assist the male in the execution of his part of the pattern. Yerkes and Elder summarised the evidence on this point in the following words:

> Our data justify the following statements of fact. Prior to sexual maturation, the female chimpanzee, if with other members of her species of comparable age and both sexes, learns from social contacts all that is necessary to enable her to behave wholly appropriately and effectively in the mating situation, and even to encourage, direct, or definitely to aid the male in copulation. By contrast, and this seems strange indeed, the recently matured male who is inexperienced in mating with a mature and receptive female commonly acts initially as does the sexually immature male, somewhat playfully, puzzled and as if at a loss how to meet the situation. Even if, with the cooperation of the female, he attempts to copulate, he usually fails. . . . Our data indicate that the role of the sexually competent male is not taken by reason of maturation solely, but instead that experience and practice are essential to biological adequacy of performance and still more to perfection and skill. Seemingly it requires months for the mating pattern of the male chimpanzee to become so far perfected that it is satisfactory to the experienced female and in highest degree efficient as reproductive function.

The zoological evidence permits two generalisations concerning the role of practice in the formation of sexual patterns. First, that the importance of individual experience tends to be greater for the higher mammals than for the lower. Second, that among the subhuman primates the male is more dependent upon specific heterosexual learning than is the female.

The question of the probable contributions of early experience to the capacity for effective coitus in the human species arises next. The evolutionary data strongly suggest that human sexual patterns are not completely organised on a strictly inherited level. Instead, it is highly probable that practice is essential to complete arousal and particularly to satisfactory expression of sexual excitement. It follows, therefore, that if they are ever to derive

maximal satisfaction from sexual relations, individuals who are reared under conditions that prevent or seriously reduce experimentation and practice during childhood and adolescence will be forced to go through the essential learning processes after adulthood has been attained.

This type of adjustment may be exceedingly difficult for young adults of either sex, particularly if they belong to a society that inculcates manifold sexual inhibitions in the developing individual. The man or woman who learned during childhood and adolescence that it was 'wrong' to examine or stimulate his or her own genitals, that it was even 'worse' to have any contact with those of another person, and, particularly, that attempts at heterosexual relations were immoral, is expected to reverse completely at least some of these attitudes on the wedding night or shortly thereafter. This expectation is difficult to fulfil. If the initial lessons have been well learned, the unlearning is bound to take a long time and may never be completed.

Even the simplest sorts of preliminary sexual instruction seem to help intelligent people to make at least partially satisfactory adjustments to the marital relationship. Dickinson compared wives who before marriage had been examined by a physician and instructed concerning the physical relations involved in intercourse, with other women who had received no such preparation. Of fifteen 'instructed' wives, 12 reported the occurrence of orgasm during marital coitus, whereas only 10 of 35 'uninstructed' women had a similar experience.

Davis asked 992 American wives if they felt they had been 'at all adequately prepared by instruction for the sex side of marriage'. Slightly more than one-half of them replied in the affirmative, but the wording of the question was so ambiguous as to result in contradictory answers. For example, one individual who said she had been well prepared added the following explanation: 'My mother taught me what to expect. The necessity of yielding to her husband's demands had been a great cross in her own life.' Others who stated that their preparation had been adequate offered comments to the effect that they had known in advance that 'men were chiefly animals' or that 'most men must be beasts'. The adequacy of instructions that produce these attitudes is, to say the least, open to debate.

Some wives answering Dr Davis's questionnaire stated that their ignorance of the facts about sexual relations had cost them dearly, and that instruction 'would have saved years of difficult adjustment'. But other women pointed out that their knowledge of the 'facts' was sufficient. The difficulty was that they 'knew nothing about *sex emotions*' which subsequently proved to be of primary importance. However, regardless of the validity of their judgements concerning the adequacy of premarital preparation, the women who said it had been adequate were, as a group, inclined to view their own marriages as relatively happy. The wives whose instruction was considered inadequate tended to describe their marital relationship as unhappy.

We have no intention of implying that the happiness or success of a marriage rests solely upon sexual factors. On the contrary, some American wives consider their marriages happy although sexual relations are described as invariably distasteful and unsatisfactory. And, contrariwise, other women find the sexual side of marriage totally satisfying but say that other elements in the relationship render their marriages unhappy. The points that seem worth making are (1) that satisfactory sex adjustment appears to increase the likelihood of successful marriage, (2) that ignorance and lack of preparation reduce the probability that such adjustment will occur, and (3) that a society which permits extensive sex play in childhood and adolescence may thereby increase the chances that sexual relations in marriage will be pleasant and mutually satisfying.

SUMMARY

In this chapter we have outlined the physiological changes that lead to achievement of full reproductive maturity. It has been shown that they are gradual rather than abrupt, and that the outward signs of puberty are indices to the onset of adolescence, which actually extends over a period of eight to ten years and precedes the life stage of maximal reproductive efficiency. Nevertheless, many human societies recognise the puberal changes as socially significant and use them as indicators of the individual's readiness to enter adult life. Among many peoples the appearance of these signs is marked by ceremonials and initiation rites.

As far as actual sexual behaviour is concerned, it develops somewhat more rapidly in certain societies than in others. Some cul-

tures fully approve of a variety of sexual practices among young boys and girls and between adolescents of both sexes. When there is any difference in treatment, the behaviour of girls is more carefully controlled than is that of boys. As long as the adult members of a society permit them to do so, immature males and females engage in practically every type of sexual behaviour found in grown men and women.

There are other societies, including that of the United States, which attempt to restrict severely the sexual activities of prepubescent and adolescent individuals. Cultural ideals may include disapproval of any sexual life whatsoever for all unmarried persons. However, even under the most severe restrictions some individuals do engage in a certain amount of erotic play.

Males and females of several sub-human primate species are known to indulge in a great deal of sexual play which leads eventually to the perfection of adult patterns of heterosexual coitus. There is some reason to believe that the practice thus acquired is essential to the performance of biologically effective intercourse in adult life. The contribution of sexual experimentation and learning in human sexual relations is unknown. But the zoological data suggest that personal experience in coitus may be very important. Such experience may be gained early in life under certain social conditions, or deferred until after marriage in other circumstances.

After reviewing the cross-species and cross-cultural evidence, we are convinced that tendencies towards sexual behaviour before maturity and even before puberty are genetically determined in many primates, including human beings. The degree to which such tendencies find overt expression is in part a function of the rules of the society in which the individual grows up, but some expression is very likely to occur under any circumstances.

Notes

1. Aleut, Andamanese, Apache, Arapaho, Arikara, Ashanti, Barama, Bena, Chagga, Chewa, Cheyenne, Chiricahua, Cree, Cuna, Dahomeans, Flathead, Fox, Havasupai, Hopi, Ila, Kiwai, Kurtatchi (nobility), Kutchin, Kutenai, Lamba, Macusi, Manus, Maricopa, Marshallese, Mataco, Nama, Nauruans, Navaho, Oto, Papago, Pima, Sinkaietk, Swazi, Thompson, Thonga, Tikopia, Toba, Tolowa, Trobrianders, Trukese, Wogeo, Yapese, Zulu. Inferred: Abipone, Aranda, Choroti, Dieri, Dusun, Goajiro, Kamilaroi, Kwakiutl, Lenge, Loyalty Islanders, Mbundu, Miriam, Nandi, Orokaiva, Palauans, Pedi, Seniang, Tehuelche, Toda, Tswana, Wolof, Yuma.

2. Ainu, Aleut, Alorese, Andamanese, Apache, Apinaye, Aranda, Arikara, Azande, Balinese, Chagga, Cherente, Chewa, Colorado, Creek, Dahomeans, Delaware, Dieri, Dusun, Gilbertese, Hopi, Ila, Jivaro, Jukun, Kababish, Kamilaroi, Kansa, Keraki, Kiwai, Kurtatchi, Kusaians, Kwoma, Lango, Lesu, Marshallese, Masai, Mbundu, Menomini, Miriam, Nama, Nandi, Orokaiva, Oto, Pedi, Ponapeans, Pukapukans, Purari, Seminole, Seniang, Sinkaietk, Swazi, Tanala, Taos, Tasmanians, Thompson, Thonga, Tikopia, Tokelauans, Tongans, Tswana, Venda, Wogeo, Wolof, Xosa, Yuma, Yungar, Zulu, Zuñi.

3. Abelam, Apinaye, Ashanti, Chagga, Chiricahua, Cuna, Dahomeans, Haitians, Kwoma, Manus, Murngin, Penobscot, Rengma, Trukese.

4. Ao (girls), Ashanti (girls), Choroti (girls), Chukchee (girls), Haitians (boys), Jukun (girls), Lamba (girls), Mataco (girls), Seniang, Siriono (girls), Swazi, Toba (girls), Tupinamba (girls).

5. Chagga, Masai, Pedi, Swazi, Thonga (inferred), Wolof.

6. Alorese (older children), Andamanese, Aranda (girls), Azande, Bena, Chagga (girls), Colorado (girls), Cree, Creek, Crow (girls), Dusun (girls), Flathead, Ganda (girls), Havasupai, Huichol, Kickapoo, Kiowa Apache, Kiwai, Klamath, Kurtatchi (girls), Kutchin, Kutenai, Kwoma, Lango, Mailu, Mandan, Mangarevans (now), Manus, Mbundu, Menomini, Omaha, Orokaiva (girls), Papago, Pedi (girls), Purari (girls), Ramkokamekra, Reddi, Rengma (girls), Seminole, Sinkaietk (girls), Tinguian, Tokelauans, Venda, Wappo, Yagua, Yako, Zulu (older children).

7. Alorese, Chewa, Copper Eskimo, Crow (boys), Easter Islanders, Hopi, Ifugao, Ila, Kazak, Kwakiutl, Lepcha, Lesu, Mangarevans (formerly), Maori, Marquesans, Marshallese, Masai, Nama, Ojibwa, Palauans, Ponapeans, Pukapukans, Samoans, Seniang, Siriono, Tikopia, Trobrianders, Walapai, Wogeo, Yapese, Yaruro, Zulu.

8. Alorese, Copper Eskimo, Cree, Dusun, Easter Islanders, Flathead, Ganda, Hopi, Hugao, Lesu, Marquesans, Ojibwa, Ponapeans, Pukapukans, Samoans, Tikopia, Tinguian, Trobrianders, Trukese, Wogeo, Yapese.

9. Ainu, Aymara, Balinese, Barama, Chewa, Copper Eskimo, Crow (boys), Dobuans, Easter Islanders, Futunans, Gilyak, Goajiro, Gond, Ifugao, Ila, Lapps, Lepcha, Lesu, Macusi, Mangarevans (formerly), Maori, Marquesans, Marshallese, Mongols, Nandi, Naskapi, Natchez, Nauruans (commoners), Palauans, Palaung, Ponapeans, Pukapukans, Seniang, Siriono (boys), Taos, Tarahumara, Thonga (girls), Toda, Tongans (boys), Trobrianders, Trukese, Tuareg, Walapai, Wogeo, Yakut, Yapese, Yaruro, Yukaghir.

Reports of children abandoned by their parents and subsequently brought up by a wild animal, usually a wolf, have cropped up from time to time. Scientific interest has, of course, centred on the degree to which such children have retained their human characteristics when reared in an entirely non-human environment. Such unusual cases should help us to judge the extent to which human behaviour is learned in a social context.

The early reports were somewhat naïve in that not enough was known about child development under normal circumstances, and some fairly rash assumptions were made about how typical the child was in the first instance before being adopted by wild animals. Later work, as represented by the following two articles, has been both better informed and more circumspect.

We are now less inclined to believe that the stories of children reared by wild animals reflect the facts. Now that we know that there is a far more likely explanation for the 'wild' behaviour, it is easier to see where the weaknesses in these earlier accounts lie.

The other important feature of the Bettelheim paper is the stress that is laid on the giving and receiving of affection from the very first, if the infant is to grow into an emotionally mature and stable person. Healthy personalities grow in healthy emotional environments. Benign social contact is essential if a recognisable human being is to emerge from the process.

THE WOLF BOY OF AGRA

William Fielding Ogburn

THE following story was recently printed in the Delhi edition of the *Times of India*:

> A six-year-old boy has been returned to his parents after being in the company of wolves for four and half years, it is learned.
>
> According to information reaching here [Agra] from a small village under the police station Khandauli an eighteen-months-old baby was carried away by wolves four and a half years ago from the house of Kanchan Jatav and all trace of the boy disappeared.
>
> The other day, a hunting party surprised a pair of wolves at the village tank. When chased the party saw a human-looking wolf accompanying the pair.
>
> The hunting party succeeded in capturing this 'wolf' and brought it to the village where he was recognised by his parents as their long-lost son.
>
> It is believed the child was reared by wolves during the period of his captivity.

The story was amplified and, no doubt, embellished in subsequent dispatches. Psychologists were reported as studying the boy. The physician who examined him at the hospital received five hundred letters of inquiry about the wolf child, which he said he had been unable to answer, since he had no secretarial help.

Being behind in my schedule, with only two articles and one chapter to write, I decided to take a few days off, go to Agra and see what I could find out about the wolf child. Besides, the temperature was only 104 degrees, and the dust storms were just beginning.

Even before Romulus and Remus, myths of human children being reared by wild animals fascinated primitive peoples. Such stories are common in India, where the climate is mild enough in winter for a child to live unclothed and where it is warm enough at night for the children to sleep outside where wolves may find them. None of these stories, however, is well authenticated, and it is not known how much is legend.

We do know that mothers of one species will suckle the young of another species. Several photographs of such adoptions were published a few years ago in *Life* magazine. One was of a scottie dog in St Petersburg, Florida, nursing young racoons. When separated from her litter of ten pups, she went into the woods and returned with three little racoons. Another photograph was of a dog in Roanoke, Virginia, feeding a six-month-old kitten a year after her last litter. She still was producing milk.

Might not a she-wolf pick up an infant left on the edge of the field while the mother was cultivating the grain and suckle it,

particularly if the wolf had recently lost her young and if her mammary glands were swelling with milk or if there were some abnormal variation in her endocrine system? We know from laboratory experiments on animals that mothers lose interest in their young offspring when manganese is removed from their diets. Perhaps mother love may likewise be accentuated chemically.

Despite popular appeal, there is little of interest to science in whether or not a carnivorous wild animal will rear a helpless human baby. We know that animals who live by killing other animals are not hungry all the time and have their playful moods. Science has had opportunities to study the rearing of racoons or kittens by dogs or of ducklings by chickens. But the endocrinologists and the neurologists do not seem to find much worth while in these phenomena.

Science, though, is greatly interested in children reared by wolves, for the light the study of them may throw on the new formulation of the old problem of heredity versus environment, which is concerned with the role of culture and of the learning process in shaping personality. Would such a child show any human traits, or would he behave like a wolf? Would the heredity of *Homo sapiens* mean that a child so reared would walk erect, talk, invent? Would he have a sense of modesty? Of shame? Would he have a soul? Or would he have no other habits than those of wolves as regards locomotion, hunting, howling, sleeping and dwelling? Or would he do these things better than wolves? Scientists would like to know how much one's personality is shaped by hereditary factors and how much by the learning process. It seemed to me worth while to take a look at this wolf boy. It was an opportunity.

So I took the ancient road which goes to Benares, the holy city of India, and passes through Agra, on which may yet be seen camels and an occasional elephant. Women in bright saris gather at roadside wells to gossip and fill their pottery vessels with water which they carry home on their heads. On this much-travelled highway, the 200-horse-powered automobiles must go off the road to pass the 2-horse-powered bullock carts, whose speed is about two miles per hour.

Arriving at Agra, I learned that the psychologist who had studied this wolf-reared youngster was on his vacation at Nainital, where one sleeps in summer under three blankets at night and wears a winter suit during the day. The physician who examined the boy was busy and could not see me until that night.

So I decided to drive the thirty miles to Ferozabad, where the soldiers are said to have delivered their find to the excellent new hospital there. But I was authoritatively told no such child had ever been admitted to the hospital.

Since there were still a few hours of hot sunshine. I drove forty miles to Khandauli, which was near the village of Nagla-jar-ka, from where the baby was lifted by the wolf. The police at the station in Khandauli told me that the baby who had been lost four and a half years had been recovered from wolves by some soldiers on a shikari and returned to its parents. They offered to send along an officer with me to the village where the boy now lived.

The villagers in Nagla-jar-ka were to celebrate a wedding; no outsiders were there. All the males, elders and children, and the little girls surrounded us as we talked, with the help of an interpreter, in the courtyard of the child's home. The women were secluded, though no doubt they listened and looked from nearby houses. It was a small village with probably not more than twenty dwellings.

Yes, the boy, when eighteen months old, was stolen four and a half years ago about ten o'clock at night as the mother and babe slept outside, on about the same spot where I was standing. The mother awoke and saw the wolf running away and the child gone. It was darkish, and the aroused family and neighbours could not find the wolf. I remarked that the wolf came rather close to the house, but I was told that sometimes a wolf actually comes inside a house. I also thought that, if the wolf went that close to a house, it might have been driven more by hunger than by pressures on the mammalian glands.

That wolves do carry away children seems to be a fact. The week I was in Agra there was a story in the newspapers about a hyena stealing a child from a village near Ahmedabad. The villagers were in a state of great fright. Another incident was told me by an Indian scientist. The child of his mother's brother was stolen by a wolf, and the theft was witnessed by the father and others. The father went in pursuit. He knew the tracks of wolves. He had seen these tracks in his morning walks to the defaecating area. Their path formed a large semicircle. He took the diameter route and lay in ambush for the wolf. As the wolf came by, he speared it, even

though it had the baby in its mouth. In killing the wolf, he hurled the spear with such vigour that he fell on the spear handle and broke it. Part of it stuck in his chest. He never recovered and died shortly after. The baby was unharmed.

To my question, 'How do you know this boy, Parasram, is your lost child?' his father pointed to a bare spot on the top of the child's head a little to the left, elliptical in shape, one and a quarter inches long by one-half an inch wide. Their baby before he was lost had a boil on his head which burst and left a scar. The scar was quite smooth, with the thick hair growing right up to the curving border of the scar. He showed me also a slightly elongated and narrow scar on the neck of the child. On this spot also their baby had had a boil.

My first sight of Parasram was as he ran to a chair, set for me, and beat on its seat with his bare hands. The noise was a bit like that of a drum. He then ran to different persons standing around in a circle and looked at them or pushed at their legs or feet. Then he went to another chair and beat on its bottom. He came to my leg and plucked at my trousers. I took him in my arms without any resistance or fear on his part, and he pulled on my nose and poked at my mouth. He soon became restless, so I sat him down. He then jumped upon the charpai, lay down, pulled up his one garment and played with his genitals. He was up again, going from one person to another. Then he laid himself on the ground and began sucking his thumb.

His mouth was open a good deal of the time. His two upper incisors were missing, but the second teeth on each side had come in, somewhat irregularly. He seemed to me small for six years. I asked my driver, who has nine children, how old he thought the child was. He guessed he was between five and six years of age and said that he was a little small for six years. The child's movements were well co-ordinated. Indeed, he seemed rather lithe in his arms and nimble on his two feet. At no time did he crawl or try to walk on all fours. The expression on his face did not appear to be dull; rather the contrary. I doubt if he smiled, yet, with his lips only partly closed, one might have thought he was smiling.

His behaviour was quite different from that described of the wolf girl Kamala, who lived in the orphanage at Midnapore, Orissa. Kamala is reported to have lain quiet, huddled in a dark corner, face to the wall. She took no interest in people or objects. At dusk she tried to run away. At night she howled. Both Kamala and Parasram ate raw meat without using their hands, and there are pictures of each eating that way.

My first impression of Parasram was that of an exhibitionistic, spoiled child whose parents have had him do his stunt before visitors who are supposed to admire him. But this child seemed not interested in the glances of the adults and appeared not to seek approval or praise. Indeed, one might think that the child was oblivious of the opinions of others. He seemed interested in people only as objects, much as one might be curious about animals in a farmyard.

The child showed no shyness while I was with him. Shyness is supposed to be related to the child's concern with himself and the reaction of others to that self. Did the child have a conception of self? I could not say. He did not talk, and knew only a few words, such as *roti*, for bread.

This boy has been moved about a good deal recently. He had been in his home with his 'parents' about a week, having come from the hospital where he was observed for a few days. He had been taken from the hospital to the home of the psychologist. A few weeks before the child had been at the house of a domestic servant, and previously he had lived four months in the family of a Brahmin. Such moving about seems not to have affected his sense of security.

The child was not afraid of a crowd. I once knew a wild American Indian who had never seen a crowd, and on seeing his first crowd he trembled, terrified.

This excessive activity of Parasram was noted by the Brahmin and at the hospital. It was interpreted as evidence of wildness. On the wildness of wild animals, I once read a book on wild wolves in Alaska, as observed by a naturalist from a distance using powerful magnifying glasses. In the day the cubs and grown wolves were not especially active. They lay around, slept, woke up and stretched. The cubs would occasionally play, much as puppies of domesticated dogs.

What the glandular balance of this hyperactive child was is not known. Of the feeble-minded I have seen in institutions, I do not now recall such excited activity.

To my question as to where they found their lost child, I was told that he was given them by a man in Ferozabad named

Rajendra Prasad Sharma, but they did not know where he lived or what he did.

In Agra a journalist who had undertaken some investigation of the capture told this story: Some soldiers, who were hunting for game, found the child in company with wolves at a stream, drinking water. The rescued child was taken to a doctor who, in turn, gave him to the police. A Brahmin named Indraj Sharma, who kept a bicycle shop in Ferozabad, having no son, then took the child and kept him for several months. He then gave him to a domestic servant of the Dharmuk caste whose home was in the village of Dadampur, four miles from Ferozabad, and whose neighbour, who was a relative of the Babulal Jatavs, told them about the boy's rescue. They came, recognised him as their lost son and took him home with them to Nagla-jar-ka.

The journalist had seen and talked with all these persons except the soldiers who rescued him and the doctor to whom the soldiers gave the child. The doctor had moved to Lucknow, where it would be difficult to find him, since he was not a trained licensed physician. He had worked in a chemist's shop in Agra. I suggested to the journalist that the soldiers who captured the boy were probably from the cantonment in Agra and that a unique experience such as this would be known to other soldiers and that the rescuer might thus be found.

The physician at the hospital, an Indian, was trained in London and was a competent observer. He said that the head of the child seemed to be a little smaller than the average and that a possible microcephaly suggested mental deficiency. In the hospital he had been restless and active. He had to be watched.

When drinking water, the child put his head down close to the water, shook his head from side to side, looking at the water before drinking. The physician said that such is the manner of the drinking of wolves that try to see that the water is clear of surface material before drinking.

The physician also brought in an Alsatian dog, a little larger than the Indian wolf. The owner stayed near, to keep the dog from biting the child should it try to do so. The child went up at once to the dog, nestled close to it and then put his hands on the dog's jaws, much as he put his hands on my nose. He even put his face up against the dog's mouth.

I suggested that this behaviour may not have indicated experience with wolves but only a manipulatory tendency common with children. The original nature of a baby is to have only two fears: of a loud noise and of falling. It has to learn to fear dogs, fur and snakes.

I next inquired if the psychologist had made a report. But the hospital authorities knew nothing of it. I wrote him, and he replied that he had not made a report but planned to do so later.

The child had lived with Indraj Sharma, the Brahmin who operated a bicycle-repair shop, for nearly four months. The next step was to talk with him if he was still in Ferozabad and if he had not gone away for a holiday.

Sharma was an athletic-looking man who had been a soldier for a decade but was now operating a bicycle shop. He was forthright, direct and apparently accustomed to giving orders, for he twice dispersed the crowd surrounding us. He got a chair for me and a charpai for himself. We sat near his shop, and he gave me a big pink drink with ice in it, the taste of which I did not recognise. He could not speak English.

The results of this interview were crucial. They gave me what I went to Agra to find out. The gist is recorded in the form of questions and answers:

Q. Shri Sharma, I would like to have you tell me who gave you the child whose parents live near Khandauli.

A. I found the child myself.

Q. I was told that some soldiers found the child.

A. No, I found him myself. I was a soldier at one time.

Q. Was anyone with you when you found him.

A. I was alone.

Q. Was he with wolves?

A. No, there were no animals of any kind near him.

Q. Was there a wolf den near him?

A. Not that I know of.

Q. I was told that a party of soldiers captured and gave him to a doctor who in turn gave him to the police.

A. He was not found by a party of soldiers. I found him myself. The child was thin, seemed not to be well. So I called in this doctor and paid him for his services.

Q. Where is the doctor now?

A. He left about two months ago for Lucknow. I don't know what he is doing or what his address is.

Q. Do you have children?

A. Yes.

Q. Did this child and your children play together?

A. No, they never did.

Q. Did the child talk?

A. No, he never talked.

Q. I heard that you gave the child to a domestic servant.

A. Well, my wife was 'fasting' and could not care for the child, so I got this woman to take care of it. I paid her two rupees.

Q. I was told the child was so wild this servant had to keep him chained.

A. She did not chain him. But sometimes I tied him with a rope, not a chain, to keep him from running away.

Q. Did he ever show affection?

A. I suppose so.

Q. How did the Babulal Jatavs learn about the child?

A. I announced over a loud speaker in Ferozabad that I had found a lost child, that I would like to know to whom he belonged and that I would return him to his parents. The child's parents heard about it and thought perhaps he was their lost child. They came and said that he was their child and took him away with them.

Q. Where did you recover this boy?

A. About two miles from Tundla.

Tundla is a railway junction about twenty-four miles by road from Nagla-jar-ka, where the child was stolen by the wolf – rather far for the wolf to carry the child or for the child to walk.

Such, then, is the account of the recovery of the child by the man who found him. His account appears to be reliable. Cross-questioning did not cause him to change. If his story is true, then the child was not captured from wolves. Indeed, the child lived with a family in Ferozabad for four months without any association with wolves being mentioned; there was no newspaper story for these four months about his being a wolf-reared boy.

The linking of this child to a wolf was made by the Babulal Jatavs when they acquired him. The Babulal Jatavs did have a child of one and a half years of age stolen by a wolf. At least, the whole community of Nagla-jar-ka, where the Babulal Jatavs live, think so. And there is good evidence that wolves do steal infants.

The false identification of the child by the Babulal Jatavs is natural. First, the child could not talk. His habits did suggest wildness. There were scars on the scalp of the child, and the lost infant had had boils on its scalp, so the parents claimed. Shrimati Babulal had had three children, the boy stolen by a wolf and two girls. The parents, like all Indian parents, craved a son. But at the time the mother had no son living, and she probably still mourned for her lost child.

The reputed father, Shri Babulal, has refused a test that might have disproved his parentage. This was the test of blood types. The physician at the hospital knows the blood type of the child, Parasram, and of the mother, Shrimati Babulal. If he could have obtained a sample of the blood of the father, he might have been able to disprove their parentage, though this test would not have proved their parentage. Why did Shri Babulal refuse? Did he have some doubt that the child was not his? Not necessarily. He would have given a sample of his blood for analysis by the physician, if the physician had paid him a large sum of money, which the doctor said was exorbitant. The father is, like Indian peasants, very poor. He expects gifts of money from those who come to see the child. At least, he is not averse to accepting money from visitors.

The child, Parasram, if not reared by wolves, was a lost child who could not talk at five or six years of age.

Children, I suppose, do get lost in India; and on the border of jungles they may not find their way home. In the nineteenth century, when famines were frequent in India and many lives were lost, the orphans were called 'famine' children and became wanderers. Not all were taken to orphanages or to relatives. It is said that village prostitutes in India abandon their unwelcome children. It may be that difficult or feeble-minded children would sometimes be abandoned, though Indians have the reputation of being devoted parents.

How long such waifs live is a question. It would have been interesting if Sharma, instead of capturing the boy, had watched him a while to see what he did and where he went. The winters in the Agra area have a temperature as low as, say, 40 degrees, which would be trying for a child to live in. Peasants in Nepal, though, go bare-footed when the temperature is below the freezing

point and when the ponds are frozen over. But they wear shawls, however, around their necks and over their heads. The monsoons of the summer would be difficult for an Indian child to live through alone in the forests. Parasram may have been in the wilds only a short time.

Then there is the problem of what such a child would eat. Berries, fruit, insects, small animals and dead ones are possibilities. It is conceivable that in some cases a stray child might live with some other animal.

Even if Parasram, 'the wolf boy', was not reared by wolves, there is still much strangeness about the child's past. If he had lived in the wilds only a short time, how did it happen that he could not talk? He was not deaf and dumb, for he learned a few words, such as terms for bread and water. If such a lost child was not by heredity feeble-minded, then a study of his personality might shed light on the roles of culture and heredity in the development of personality. It was not possible for me to remain longer to study Parasram, as my tickets for an early departure from India were already bought. One wonders if a child psychologist, observing a waif like Parasram without knowing his past history, could reconstruct the early socio-psychological experiences that so shaped his personality.

While my hunt for a wolf boy was not successful, as a little journey into a land where myths originate, it was successful in showing how a myth began.

It is not to be inferred from this account that children may not have been reared by wolves. They may have been. But I found no evidence that this 'wolf boy' ever saw a wolf.

FERAL CHILDREN AND AUTISTIC CHILDREN

Bruno Bettelheim

In science more than in other fields of human endeavour, the correction of a widely held error often contributes more to the solution of a thorny problem than some new discovery or theory. Often, too, erroneous ideas can prevent the valid knowledge we already have from making its influence fully felt. This is amply demonstrated by Professor Ogburn's 'The Wolf Boy of Agra'. For years, on the basis of much experience with severely autistic children, I have been convinced that most of the so-called feral children were actually children suffering from the severest form of infantile autism, while some of them were feeble-minded, as was possibly the Wild Boy of Aveyron.[1]

Children suffering from early infantile autism typically are unable to relate themselves in the ordinary way to people and situations from the beginning of life. Extreme aloneness shuts out anything that comes to them from the outside. Some acquire the ability to speak, while others remain mute. But language is not used to convey meaning to others. The overall behaviour is governed by an anxiously obsessive desire to maintain sameness. The term 'feral child', in contrast, is not a definite diagnostic category but vaguely denotes very wild children and those supposedly reared by animals.

From historical accounts of most of the feral children, diagnosis cannot be established. But, the more detailed the accounts, the more definitely do they seem to signalise autistic children. Fortunately, in the case of the two famous wolf girls of Midnapore, Amala and Kamala, a fairly accurate description of the behaviour of the older girl and of the steps in her partial recovery has been published.[1] This story closely parallels our experiences with autistic children at the Sonia Shankman Orthogenic School – a laboratory school of the University of Chicago devoted to the education and treatment of severely disturbed children. These children have never lived in the company of wolves and were reared by none but human beings. The similarities were so great and so unmistakable that no other conclusion seemed possible but that the two wolf girls also had suffered from severe infantile autism, accountable without a history of being reared by animals. Yet I could not doubt the veracity of the Reverend Singh's description of this kind of behaviour and development, so familiar to all of us. In fact, I was probably more ready than most to believe in the accuracy of his account, and therefore I fell into the error of also giving credence to his report of how he found them.

But, as I read Ogburn's story, suddenly the blinders fell from my eyes. Now it became all too clear: Singh's account of his close association with the girls could be, and, I believe, was, entirely correct; his *interpretation* of the origin of their behaviour – their having been reared by wolves – was false. He was carried

away by his imagination about that one single event which makes or breaks his interpretation, namely, the way in which the children were found. He flushed three grown wolves out of the hollow of a 'white-ant mound as high as a two-storied building' and found 'two cubs and the other two hideous beings there in one corner, all four clutching together'.[2]

Hindsight is always easy. I can now also see the parallel between Singh's story of finding the girls and the wild fantasies we spun about the pasts of our autistic children when we first met them – speculation originating in our efforts to find emotionally acceptable explanations for this nearly inexplicable and wholly unacceptable behaviour.

Later we came to recognise that this speculation originated mainly in two different psychological needs: first, in our narcissistic unwillingness to admit that these animal-like creatures could have had pasts at all similar to ours (the same narcissism which revolted against the theory of evolution) and, second, in our need to understand and to explain (the more unusual, the less acceptable a phenomenon, the greater our narcissistic need of an explanation). Perhaps, too, the more revolting the behaviour, the less we wish to devote much thought to it, explaining it by an emotional reaction or in such simple form as to require no further thought.[3]

Throughout the story of Amala and Kamala, the strangeness of the children's behaviour and that Singh had proved truthful in his account of it were taken as evidence that he must also have told the truth in all else: in his story of the way he found them and in his interpretations of how their correctly observed behaviour was due to, and could only be explained by, their having lived with wolves. The mechanism at work here seems to be that the rational mind, which at first rejected the story of Kamala's behaviour, turned out to be an unreliable instrument. Therefore its critical voice was silenced, as far as these stories were concerned, and henceforth everything was believed as told.

This is not surprising to us. Many times when we have described the behaviour of some of our extremely autistic children – how they urinated and defaecated without so much as knowing it as they walked or ran about; how they could not bear clothes but would run about naked; how they did not talk but could only scream and howl; how they ate only raw food; how they would bite us so often and so severely as to require frequent medical treatment – even persons quite familiar with disturbed children would react with polite or not so polite disbelief. But later, when they met these children, their doubts changed to complete belief, so that they would have been willing to believe almost anything told them about the children or their pasts.

Thanks to Ogburn's investigation, we now know that Parasram was not found in the company of wolves. We have, therefore, good reason to doubt that Amala and Kamala were found living with wolves. But, just as it was true that Parasram was found wild, there seems no reason to doubt that Amala and Kamala were found living wild in the forest.

But how did these children survive all alone in the wilderness? How did they get lost in the first place? I believe from our experience with autistic children that the wild children could not have survived for very long by themselves, even allowing for the clemency of the Indian weather. Neither their haggard look nor the absence of clothing nor the 'hideous ball of matted hair' prove that they had been lost for long: some of our autistic children keep their wild looks for months. They can and do tear off all their clothes in minutes. Even after years with us the well-groomed hair of one of our autistic girls could, within hours, turn into a 'hideous ball of matted hair', glued into a mass by saliva, remnants of food, dirt and whatnot. One of our autistic girls kept her face well hidden for months behind a curtain of hair.

Ogburn speculated that the Indian children might be lost, feeble-minded or the abandoned offspring of prostitutes – these offer as accurate explanations as any other. My guess would be that they were simply emotionally, and perhaps also physically, abandoned. In the report on the two children of Midnapore we are told by Bishop H. Pakenham-Walsh: 'The very primitive people who inhabit the parts where these children were found, who are not Bengalis, do fairly frequently expose baby children.' If these people are ready to expose normal babies, is it so far-fetched to believe that they may also expose older children who act like babies (as autistic children do) or children who seem extremely abnormal to them? Also, how old are 'baby children'? How old are the children before people normally stop exposing them? All these questions remain unanswered in the account of Amala and Kamala.[4]

Our own experience suggests the explanation that the girls in

question were probably utterly unacceptable to their parents for one reason or another. This is characteristic of all autistic children, no matter of what age; the parents manage to disengage themselves from them by placing them in an institution (as is the usual case in the United States today), or by setting them out to fend for themselves in the wilderness, or, the most likely explanation, by not pursuing when they run away.

Our experience with the parents of autistic children, many of whom are well-educated, good, middle-class people, leaves little doubt that in their deepest emotions they wished to be rid of them and for very good reasons. They could not afford to become conscious of such wishes or to act upon them because of the demands of conscience, the behaviour expected of parents in the United States and the near-impossibility of a child's getting permanently lost in our cities. But there is hardly one such parent, living in twentieth-century America and conscientiously on the watch, whose autistic child has not managed to get lost several times after the age of about three or four.

It is reasonable, therefore, to assume that it must be quite easy for such children to get separated from their parents under more primitive conditions of life. If one can draw any conclusions from our experiences with some of the parents of autistic children, it is reasonable to assume that their efforts to find their lost children have been more than lax. So much for how feral children may get lost.

But what promotes the belief that there are feral children in general or wolf children in particular? First, such children are not dumb but do not talk; and speech, almost more than anything, separates humans from animals. Animals cannot talk; hence these children must have something in common with animals. Second, all normal children, even if feeble-minded, need humans to take care of them and will reach out to them; but these children shun human company. Third, some of these children are ferocious in their attacks on others, using claws and teeth, like animals. Beyond that, I can offer only speculation. Thus, for example, if the people of the region where the two girls of Midnapore were found believed in the transmigration of souls and were confronted with the behaviour of the girls, is it not possible that they thought of them as having been wolves in a previous incarnation or of now representing an incarnation that was part wolf, part human? Such beliefs were unacceptable to the Reverend Singh, but perhaps they were part of his ingrained thinking before he became a student of Bishop's College in Calcutta.

I have already suggested the role that human narcissism may have played in making credible the stories of feral children. As long as we could believe that insane persons were possessed by ghosts or devils, their 'wildness' was less of a shock to our self-image as human beings. But the origins of the sub-human, animal-like behaviour of these children are, in our enlightened age, no longer sought in the world of spirits. In this day of reason, we think, of course, of childhood environment as the source of their behaviour. But on first encounter with their wildness, and, thereafter, when their total withdrawal, their 'contrariness', their violence or other types of inhuman, animal-like behaviour have overpowered us and made it harder for us to deal rationally with their onslaught, then, and despite all our knowledge, we too are thrown back for moments to feel that they are possessed – that they are 'animals'. To quote a typical response after a rather mild example of such behaviour shown by Anna, one of our wild children: 'As I watched her continual application of saliva to all parts of her body, her biting and chewing of her toes, I thought to myself, "She is an animal, destructively washing herself."' So the easiest solution to the problem of their behaviour is to believe it the result of an animal upbringing.

There are other and more specific reasons to suggest comparing these children with animals. During one year a single staff member had to have medical help more than a dozen times for bites she suffered from Anna, and the children regularly bare their teeth when annoyed or angry. Different, and again reminiscent of animals, is their prowling around at night, in marked contrast to their quiet withdrawal into a corner during the day. One of these girls could finally be reached and brought to accept closeness only when her favourite counsellor roamed through the building with her for many hours during the night. And only then would she accept food from us. Then there is their great preference for raw food, particularly raw vegetables. Some will go to almost any length to get raw onions and lettuce and similar food, and go into violent temper tantrums if they do not get them immediately. Others lick salt for hours, but only from their own hands. Others again build themselves dens in dark corners or closets, sleep no-

where else and prefer spending all day and all night there. Some build caves out of blankets, mattresses or other suitable objects. They do not permit us to touch either them or their abodes, and at least two of them would eat only if they could first carry their food into their self-created caves or dens, where they would then eat without using utensils.

Some of these children, on seeing animals, respond as though they had found a dear, long-lost friend. One girl, for example, became extremely excited on seeing a dog; she showed a strong desire to run towards it and cried or howled like an animal, particularly like a wolf. She fell on all fours, jumped like a dog with her head down, and made biting gestures. Now, had we believed in the feral origin of this girl – whose total life-history, incidentally, is well known to us – we would probably have been convinced that, on seeing that wolf-like creature, she was filled with memories of her happy times among wolves and was reverting to what she had learned from them.

And yet, despite such similarities between the behaviour of some of our autistic children and that of the wild children described in the literature, could it not be that this is all due to chance – that the similarities are only superficial and that closer inspection would reveal important differences? To decide whether Amala and Kamala were autistic children, we have to ask whether they showed all characteristics typical of the disturbance. Does what was viewed as the characteristic behaviour of these two girls resemble the characteristic behaviour of our autistic children and of infantile autism, as reported in the literature?

Singh, in describing the children's behaviour, states as most typical of the two wolf girls, first, what he calls their aloofness and, second, their shyness or fright.

> The presence of others in the room prevented them from doing anything, even moving the head from one direction to the other, or moving about a little, changing sides, or turning about. Even a look towards them was objectionable. They wanted to be all by themselves, and they shunned human society altogether. If we approached them, they made faces and sometimes showed their teeth, as if unwilling to permit our touch or company. This was noticed at all times, even at night. . . . For nearly three months . . . there was a complete dis-association and dislike, not only for us, but for their abode among us, for movement and play – in short for everything human. [*Sic.*]

A child psychoanalyst described as what he considered most characteristic of one of our ten-year-old autistic girls: 'Her most pervasive behaviour trait is her overwhelming panic reaction to the slightest interference from the outside, often occurring without visible external motivation.'

This is behaviour typical for all autistic children. But how do the authors explain it? According to them, withdrawal from the surrounding world was related to the following:

> After their rescue and subsequent capture they were looking for the cubs and the wolves. It was noticeable that they wanted their company and association, but finding that they could not get them here, they refused to mix with the children or with anybody. . . .
>
> They could not find their mates in the jungle; they could not prowl about with the wolves; they missed their cozy den, and could not get to feed on meat or milk. Consequently, the thought of their old environment preyed heavily on their mind, and their thought was to regain their former habitation and company. This fact made them meditative and morose.

But how could it have been 'noticeable that they wanted' the company of wolves, or how could it have been deduced that 'the thought of their old environment preyed heavily on their mind' and that 'they missed their cosy den'? Formerly autistic children whom we have succeeded in rehabilitating to the point where they could tell us about their autistic past have had only the dimmest notion of the vague fantasies which occupied their minds in a state of total withdrawal. We can say with certainty that those who were mute for long periods and were later able to reflect upon it recalled only having been in vague states of terror, interrupted by equally vague fantasies of a bliss without content which, in the literature, has been discussed as oral reunion fantasies. But the latter gives a much too specific content to what are either vague and basically empty states of withdrawal from content and relations to environment or intervals of relative comfort and discomfort.

Singh's descriptions – trustworthy as they are – are so flavoured

by his convictions about the girls' feral past that I would like, before comparing them with our own experience and with the literature, to quote the only other eyewitness account of Kamala's behaviour because it is more concise and less adumbrated by speculations. We owe an excellent description of Kamala's behaviour when she was estimated to be fourteen years old (some six years after she had been found) to Bishop Pakenham-Walsh. He was, apparently, if not also the most intelligent, certainly the most highly educated person who saw her. He reports:

> When I saw Kamala, she could speak, quite clearly and distinctly, about thirty words; when told to say what a certain object was, she would name it, but she never used her words in a spontaneous way. She would never, for instance, ask for anything she wanted by naming it, but would quietly wait till Mrs Singh asked her, one by one, whether it was so and so she wanted, and when the right thing was named she would nod. She had a very sweet smile when spoken to, but immediately afterwards her face resumed an appearance of unintelligence; and if she were left alone, she would retire to the darkest corner, crouch down, and remain with her face to the wall absolutely listless and with a perfectly blank expression on her face. She had an affection for Mrs Singh, and was most amenable to her directions during the time I saw her. She was not interested in anything, nor afraid of anything, and cared nothing for the other children, nor for their games. She walked upright, but could not run.
>
> I saw her again two years later (when her age would have been about sixteen) and except that she had learned a good many more words, I did not notice any mental change.

During the past years we have cared for at least nineteen children at the Orthogenic School whose diagnosis was definitely infantile autism. We have lived with each for at least one year, and with most of them for several. If an intelligent and interested layman such as Pakenham-Walsh had observed most of them during their first year with us (and a few even during their second or third year) for a period of time comparable to that he spent with Kamala, his description might have been exactly like that he gave of this wolf girl. At present, we have twelve such children, ten of whom, for a year or two, showed the same behaviour, though most have since advanced much farther intellectually. Three who have been with us for less than two years are now at the stage of behaviour described by Pakenham-Walsh as seen in Kamala at about fourteen; a fourth has not yet reached it, since he has not yet said a single word. He has been with us for about a year.

Interestingly enough, two of the children began to say their first word after about a year with us, which closely resembles what was reported of Kamala – that she said her first words after thirteen months, when she began to 'prattle like a baby'. From our experience, as well as from Pakenham-Walsh's description, I doubt that she prattled like a baby, for that is the result of a voluntary reaching-out and an enjoyment first of vocalisation and later on of verbalisation. These autistic children, even after they have acquired the ability to say a few words – or perhaps, I should say, have overcome their reluctance enough to venture a few words – prefer not to use them but to let us do the talking for them, as was noted in Kamala after six years. The soft, hesitating, often echolalia-like saying of preferably only short single words which characterises their speech is very different from what develops out of the usual happy prattling of babies. It is also in stark contrast to their wild, ear-piercing screaming, which makes their barely audible and unclear (or overclear) strained enunciation of single words appear, by contrast, to be even more minimally invested with the positive wish to talk than might otherwise seem true.

Kamala's behaviour, as described, is also typical of the behaviour of children suffering from infantile autism as reported in the literature. In their classical descriptions of this disease, Kanner and Mahler state the most characteristic feature of infantile autism to be a profound withdrawal from contact with people, an obsessive desire for the preservation of sameness that none but the child himself may disrupt on rare occasions, and this while retaining an intelligent and pensive physiognomy. Equally characteristic is either mutism or the kind of language apparently not intended as communication. The children are unable to relate themselves to people and situations from the beginning of life, are referred to as self-sufficient, act as though people are not there and give the impression of silent wisdom.

Kanner's first case, Donald, at the age of five displayed 'an abstraction of mind which made him perfectly oblivious to

everything about him. He appears to be always thinking and thinking, and to get his attention almost requires one to break down a mental barrier between his inner consciousness and the outside world.' This sounds like Kamala's crouching in a corner for hours, as if meditating on some great problem, so indifferent to all that was going on that her attention could not be drawn to anything.

Of Kanner's 'Case 9' we are told that 'the most impressive thing is his detachment and his inaccessibility. He walks as if he is in a shadow, lives in a world of his own where he cannot be reached. No sense of relationship to persons. He went through a period of quoting another person; never offers anything himself. His entire conversation is a replica of whatever has been said to him.' The similarity to the wolf girls' aloofness and shyness is striking, and this boy's never talking spontaneously approximates the Pakenham-Walsh description of Kamala.

And yet, perhaps, some of the specifics of Kamala's behaviour (Amala was so much younger, and died so early, that very little of Singh's account deals specifically with her) are so different from the behaviour of autistic children as to justify Singh's belief in its feral etiology. To find out whether this is so, I have tried to make a content analysis of the wolf children's behaviour. Now, among Singh's descriptions, only one item stands out as very strange, and it is the one repeatedly referred to as explicable only by feral experience. It is also the only one for which we have no parallel among our autistic children. This is Amala's and Kamala's inability to walk erect when they were first found. While several of our autistic children have preferred to crawl on all fours for some time, and others for a long time would walk only bent over, none was actually unable to walk erect when we first met them.

Nevertheless, some facts in Singh's account of the children's lives might suffice to explain this phenomenon. On their capture the girls were immediately placed in 'a barricade made of long poles, not permitting the inmates to come out. The area of the barricade was eight feet by eight feet.' Singh left them there in that narrow confine and returned five days later to find them deserted by their keepers and left without food or drink. 'The situation [was] very grave, . . . the children lying in their own mess, panting for breath through hunger, thirst and fright. . . . The feeding was a problem. They would not receive anything into their mouths.' Finally, he got them to suck some tea from a wick, like babies. Before the girls had time to recover, a journey of seventy-five miles was undertaken, lasting for seven days, during which they were transported in a jolting bullock cart. There they spent another seven or eight days in narrow confinement. When they arrived at Midnapore, 'they were so weak and emaciated that they could not move about'.

If this had happened to some of our autistic children, we would assume that complete and prolonged deprivation of food, drink and a chance to move about are sufficient to explain total regression to infantilism, such as not walking and being able only to suck. Most of the descriptions we have of the Indian girls' way of walking refer to the time after their arrival at Midnapore. In an entry dated twenty days after (i.e. 24 November 1920), Singh mentions 'extensive corns on the knee and on the palm of the hand near the wrist which had developed from walking on all fours'. The sores healed, but, he adds, only 'on the nineteenth of December we found them able to move about a little, crawling on feet and hands'.

Thus no moving about was actually observed in Amala and Kamala after their capture, and probably none took place from the day of their capture until some sixty-two days later when they began to crawl like normal infants: a type of behaviour which can be fully explained by the deep regression they had experienced in all other ways.

The belief that their walking on all fours was due to their living with the wolves is pure conjecture.[5] I believe that it is much more reasonable to assume that their walking on all fours was part of a regression to the crawling stage such as we see frequently in some of our autistic children.

A characteristic of the children attributed to the feral past is that their eyes were wide open at night, like those of a cat or a dog. We are also told that they could see better by night than by day, though no objective tests support the assertion. This unusual acuity of vision is reported as of 20 December, the day after they first began to crawl, when they were barely emerging from their deepest immobility and debility. How Singh could be sure that they could see better at night than during the day, at a time when they could hardly move or do anything else, is beyond me.

Another capacity linked with their feral past is their ability to

smell meat or anything else from a great distance, like animals. Hardly anyone who has worked with psychotic children and has reported on them in any detail has failed to remark on their strange hypersensitivity to sensations of smell and touch in stark contrast to prolonged periods of unresponsiveness to those of sight. Hearing often takes a middle position, being sometimes blocked out and at other times or in other cases increased. In general, the senses of closeness (touch and smell) and distance (hearing and seeing) are invested in psychotic children inversely to what is usual in normal persons. Elsewhere I have reported on the extremely acute sensitivity to smell of schizophrenic children, who could smell what we could not.[6]

Kamala's ability to find her way about in the darkness is reported as unusual and likewise due to her feral experiences. But this is not unusual for many of our autistic children, who, in general, rely very little on sight for getting about.

Amala and Kamala are reported to have eaten and drunk 'like dogs from the plate, lowering their mouths down to the plate', to which statement a footnote is attached, saying: 'Their methods of eating were conditioned reflexes learned from the wolves.' Joe (whose history is given below) has never to our knowledge eaten any other way and still eats only thus, after more than a year with us. Other autistic children eat only by shovelling food into their mouths with a paw-like motion, while again others feed only from their own skins.

We are told that 'the perception of cold or heat was unknown to them', to which, in a footnote, is added that this was 'another conditioned reflex from their experience with wolves'. But some of our autistic children have sometimes tried to run out into the street stark naked even in Chicago's winter weather, when the temperature is quite different from that of Midnapore. We always caught them quickly, yet they seemed totally insensitive to such experiences and never had so much as a cold in consequence. Schizophrenic children often behave as though they were totally insensitive to heat and cold, as did Amala and Kamala, whose attitude to temperature is therefore hardly unique and proves nothing about a feral past.

Sensitivity to pain in psychotic children is, in the same way, unlike that of normal children and has to do with the nature of the disturbance and not with a feral past.[7]

This leaves only one inhuman characteristic in my catalogue – their inability to laugh. This, too, is quite characteristic for most, if not all, autistic children. As far as my recollection goes, our autistic children have never laughed until such time as we believed they had definitely moved from infantile autism towards severe neurosis, a much more benign degree of emotional disturbance.

Actually, the catalogue of animal-like behaviour is slight. Comparing it with what animal psychologists, such as Lorenz, can tell us about the wide variety of an animal's behaviour, not to mention the incredible variety of human behaviour, it appears that even our most autistic children show only a few characteristics that lend themselves to comparisons with animals. But the few characteristics or types of behaviour are so shocking to us that they assume importance and proportion entirely out of line with their actual frequency or significance in the child's total life. If we were to catalogue the behaviour of even wildly acting-out autistic children, two things would stand out: first, that most of the time they do nothing and avoid any contact with the world (cf. the aloofness of the wolf children) and, second, that if they cease to be aloof for short periods, even then they do very few things in comparison with normal children of their age. When they occasionally engage in animal-like behaviour, we tend to be so overimpressed by it that we lose sight of the rest.

What was the actual nature of the children's past? If feral experiences do not explain their behaviour, what experiences do? I can say nothing about the past of the two wolf girls, but perhaps the past of our autistic children may suggest what their actual, as opposed to their imagined, past may have been. I would like to stress here that, while we continue to study these children intensively, as of the moment we reserve judgement about what *causes* infantile autism. By now we are pretty sure of the important role of certain contributing factors, but whether they are causative or only contributory only further work will tell.

It should be stressed again that only a small minority of known autistic children are 'wild'. For example, only a few of Kanner's cases showed some of the traits which characterised Kamala and possibly also Amala; most of his cases were much more like Parasram. We have worked with both those similar to Parasram and those more like Kamala; both groups have all the essential

features in common, with the one exception of the animal-like wildness of the latter.

What causes this difference in behaviour? For a time we thought it might originate in the difference in home backgrounds. All Kanner's cases are children of highly intelligent parents, and that is also true of a majority of our nineteen autistic children. Some of our wildest children had unusual experiences in infancy, as will be seen from the life-history of Anna. Her parents would hardly be described as highly intellectual. Two other very autistic children come from highly intellectual homes. On the other hand, at least five of our quieter autistic children who are very like those described by Kanner come from non-intellectual homes, and four more from lower-middle-class backgrounds, the parents having had little, or at best modest, education.

On the quiet autistic children, more similar to Parasram, ample material is available, particularly in the case studies published by Kanner. I shall give the life-history of Anna, a wild child with animal-like traits, of whom I have already spoken:

> Anna came to us at the age of about ten. For years, before she entered the Orthogenic School, her uncontrollable wild behaviour had made life unbearable for her family. Her brother, six years younger than she, had been in constant danger of his life and had to be protected at all times from her violence. Neighbours had had to call the police because Anna was so dangerous to their children. Several efforts at placing her in treatment institutions failed. In a very well-known institution for disturbed children she lasted barely half a day; in those few hours she managed to throw the institution into such a turmoil, and did so much damage, that she could not be kept. Even in a psychiatric hospital she could remain only a month because it, too, was not equipped to handle such a wild child. There she had to be in a maximum-security room, that is, a room without any furniture, where she spent her days naked because she tore off all clothes that were put on her. Most of the time she crouched in a corner in total withdrawal; from this she emerged for short periods of wild screaming, running, jumping and pounding on walls and door. Since this made it impossible to keep her in the children's ward, the hospital had to place her in the adult maximum-security quarters, an arrangement too unsuitable to be continued.
>
> Anna's life began in a dugout under a farmer's house in Poland, where her Jewish parents were hiding from the Germans, who were trying to exterminate all Jews. Her parents were ill mated. The mother, who found the father utterly unattractive, had rejected him for years while he courted her hopelessly. Both felt that they were of unlike temperaments and background. By the time World War II broke out, the father had given up hope of winning the mother, but the German invasion of Poland soon changed the situation. He foresaw what would happen after Germany occupied Poland, and so he collected a large amount of wool and made arrangements with a gentile peasant friend to stable it in a dugout under his farmhouse, where he had set up a loom. When the Germans began to exterminate the Jews, Anna's father took permanent refuge in his small earthen cellar. But first he tried once more to persuade the woman he loved to join him. This proposal she again rejected without hesitation. She had no use for him, she said, and would rather be killed by the Germans than live with him. Soon things grew much worse; most of her family was killed. At that time the father, who could no longer leave his hiding place, again sent word to her through his gentile friend, asking her to join him. By then she had been left all alone and had no place in which to hide from the Germans. Very much against her will she took refuge with the father in his hole under the ground; his peasant friend was willing to let them both hide there. But her condition for accepting was that they would have no sexual relations.
>
> The father managed to support them, and in part also the peasant who hid them all during the German occupation, by weaving in his underground hole. The peasant sold the sweaters he wove, and on what he got for them (clothing being at a premium) he and the two in hiding were able to live. But the dugout was so small that there was not enough space for the parents to so much as stretch out at night unless the loom was taken down. Then they could bed themselves down for the night, the wool serving as bed and cover. So every night the loom was taken apart, and every morning it was re-assembled. Several times the Germans searched the farmhouse but did

not find the two in their cellar; its trap door was covered with stamped earth, like the rest of the farmhouse floor.

At least once (according to other stories they told us, it was *several* times) the Germans shot into the farmhouse. With the passing of time, life under such conditions became ever more difficult, the two being forced on each other without respite. Nevertheless, for over a year Anna's mother refused to live with her husband as man and wife. She rejected him because she felt him to be beneath her, culturally and socially, and repulsive physically. According to Anna's father, though incensed at the continued rejection, he respected her wishes and did not force himself upon her.

About what happened then, the parents' stories disagree. According to the father, they trembled for their lives every day, but he at least had his work to keep him going, while Anna's mother was beginning to lose all will to live. In desperation he decided that, if she had a child, it would restore her wish to live and maybe even make her accept him. So he convinced her to have a child, and she agreed to have sexual relations just for that purpose. Only because of these circumstances did she become pregnant.

According to the mother, the father pursued her sexually all along. After a year, no longer willing or able to stand the presence of a woman whom he loved so much and who rejected him, he threatened to kick her out of their refuge: she must either surrender or leave – which was tantamount to being killed by the Germans. Only under such duress did she finally give in.

When, in the spring of 1943, Anna, the child of this relation, was born, she did, it is true, occupy the mother and give her some interest in life, but it made existence even more difficult in their narrow confinement. When Anna tried to cry, as infants do, one of the parents had to hold a hand over her mouth, since any noise, particularly a baby's crying, would have given them all away. Also the peasant, who with reason feared for his life if it should be learned that he was hiding Jews, became more and more fearful and angry when the infant made any noise or otherwise complicated matters. So the parents and the farmer, each afraid of the Germans, did their best to see to it that the infant was totally quiet at all times and as little of a bother in all other respects.

As long as the mother could nurse her, Anna had at least enough food. But her milk gave out before Anna was a year and a half old. Then all she could feed her was raw vegetables or such like, since they could not cook. Not until 1945, when the Russian occupation replaced that of the Germans, did things improve, but by then Anna had become unmanageable. Nightly she would run, jump up and down and scream, sometimes for hours, sometimes all night. She never fell asleep before two or three in the morning. When she was not screaming or being violent, she was doing nothing, 'thinking, thinking, sitting by herself and thinking her own life'.

Eventually, the parents managed to reach Germany and entered first one, then another, and finally a third D.P. camp. But, once in Germany and relative freedom, the mother began illicit relations. When her husband learned of it, new and violent fights broke out between the two. The mother wanted to leave him once and for all, but Anna stood in the way. The mother wanted to keep the child, but her lover did not want Anna. The mother was ready to give Anna up because she wanted to live with her lover but was unwilling to surrender Anna to the father. So she suggested that Anna should stay with her own mother. To this the father would not agree – he wished to emigrate to the United States, where he had relatives, and to take Anna there.

During the years in Germany the parents frequently contemplated divorce, but at the last moment the father could never consent to it, fearing that Anna would probably be given into the custody of the mother, who had no use either for Anna or for him. There were violent outbursts in front of Anna. As one of them put it: 'We screamed and fought all the time over the child.' The father's feeling about his wife can best be expressed by his lament: 'I so often gave my life for her, and she only betrayed me.'

Long before Anna came to this country, even before her brother was born, she was examined by an American physician in one of the German D.P. camps and immediately recognised as an autistic child who needed treatment in an institution. Since we are here concerned with the background of so-called feral children, and since Anna was recognisably both wild and autistic when about

five or six years old, nothing more needs to be said here of her story.

We can say with conviction that it is simply due to chance that our two wildest girls were foreign-born and first saw the world in time of war. Thousands of children were born in D.P. camps and developed normally, and most of our autistic children were reared in what seemed like good middle-class homes. Even more than the deep inner rejection, total emotional isolation makes for autistic withdrawal, though, as said before, we reserve final judgement until we know more, In general, they seem potentially very bright and very sensitive children. Perhaps this is why they react so strongly to emotions in their parents which they somehow comprehend as a threat to their existence. To protect their lives, they stop existing as human beings, or so it would seem. Thus to give the story of one of our wild, autistic children from a middle-class family:

> Joe was the native-born son of native-born, highly intelligent and ambitious parents, exactly what Kanner describes as typical of all autistic children. He was nine when we first met his parents. By then, what they recalled of their attitudes towards him as an infant were highly coloured by feelings of guilt, but clearly this was a case of extreme neglect and isolation.
>
> The earliest investigation of Joe's past took place when he was not yet three years old. Several more psychiatric studies preceded his coming to the Orthogenic School and our interview with the parents. From each emerged the same picture of Joe's early life. Both parents began psychotherapy a few years after Joe's birth, and they gave us permission to confer with their therapists, who stated that they each gave us a truthful account of their early handling of Joe – accounts which also tallied with their personalities and with their past and present attitudes as these were made plain during treatment.
>
> Joe was born within ten months of the parents' marriage, a time when both parents were overburdened and physically and emotionally exhausted. The father, a junior in medical school, held two jobs, one being night work, to support himself and his family. Understandably, he was always on edge. The task of taking care of a baby frightened both parents. As was typical of the father all his life, when in fright, he attacked. The mother's reaction to having a baby was one of fear and panic, which only increased the father's anger. What the father called his anger and fighting back at the baby are described by the mother as violent rages, which kept her in constant fear. Finally, as she said, after living 'in fright and trembling', she suddenly turned and started a 'counter-offensive'; her husband appeared simply as an enemy to be vanquished.
>
> While the mother stated that she was 'thrilled at the idea of having a child', the father reports that her attitude changed immediately after Joe's birth. She became depressed and developed great fear, if not panic, at nursing him. Indeed, she became afraid of everything about Joe, particularly about whether he would get enough to eat. At the same time she was worried by sore nipples and confused about how often she should feed him.
>
> Joe was obviously not a happy baby. He rocked a great deal, scratched his face severely and cried a lot. He was colicky, and by the end of his first month of life both parents were 'fed up with him'. They accepted a paediatrician's advice to leave him strictly alone, particularly when he cried. The mother, who had previously felt that Joe's demands on her were monstrous in their excess, was glad to follow this advice rigidly. After a few weeks his prolonged daily crying spells stopped, but he was still left alone most of the time. His mother related, for example, that, when Joe was about six months old, 'we again had a violent quarrel one day. We screamed and physically fought each other for half an hour, or longer. Before the fight started, I had just put Joe on the potty, and it all took place within his hearing. He just sat there on the potty without moving or any reaction.'
>
> When he was not yet a year and a half old, the mother went to the hospital to avoid a miscarriage, and, since the father was a physician, it seemed simplest to place Joe in the paediatric ward. This precipitated a regressive episode in which he resumed persistent thumb-sucking and rocking and ceased to speak the few words he had already learned. Some weeks later the mother aborted and had to be hospitalised for a time; yet, despite the bad effect they knew it had on Joe, they put him again in the paediatric ward for the sake of convenience. By then both parents had lost interest in him. The father withdrew entirely

into his work, the mother became engrossed in a new pregnancy, and Joe spent most of his time alone, either in the yard or at a nearby beach. He had nobody to play with; he did not move about; he spent all his waking day simply clinging to one toy or another, such as a ball.

The parents first became aware of the seriousness of his difficulties when he was two and a half, at which time the birth of a brother intensified his symptoms: twirling, rocking, thumb-sucking and lack of speech. When they tried sending him to a nursery school, his total withdrawal became even more apparent. Treatment was attempted but failed.

The parents wished to believe his difficulties were organic, but complete physical examinations at three outstanding medical centres revealed no supporting evidence: each time the conclusion was that his difficulties were emotional in origin. The findings agreed on Joe's extreme intellectual retardation and the severity of his emotional disturbance, as evidenced by his total withdrawal and self-preoccupation, his inability to relate himself emotionally to others or to make any meaningful contacts at all, despite attempts at physical contact made by others. He was and remained withdrawn in his own autistic world, and there was no tangible evidence of fantasy content to his solitary infantile play or to his primitive hand and mouth activity. The low affect responses and the emptiness of his emotional and intellectual life indicated a primary psychotic disorder in which at no time was there any but the earliest and most primitive ego development.

The diagnostic impression was, once, of psychosis of childhood; twice, of infantile autism. Treatment away from home was recommended, and Joe, not yet four, was placed in an institution where he remained without much change until he was about nine and entered the Orthogenic School. By then he had never made any articulate sounds, though he understood simple commands. He tore his food with his hands, licked the plate like a dog, attacked others in all ways, including clawing and biting – in short, behaved like a 'feral' child.

To sum up: Study of the so-called feral children, and comparison of them with known and well-observed wild autistic children, suggests strongly that their behaviour is due in large part, if not entirely, to extreme emotional isolation combined with experiences which they interpreted as threatening them with utter destruction. It seems to be the result of some persons' – usually their parents' – inhumanity and not the result, as was assumed, of animals' – particularly wolves' – humanity. To put it differently, feral children seem to be produced not when wolves behave like mothers but when mothers behave like non-humans. The conclusion tentatively forced on us is that, while there are no feral children, there are some very rare examples of feral mothers, of human beings who become feral to one of their children.

Notes

1. J. M. C. Itard, *The Wild Boy of Aveyron* (New York, 1932); J. A. L. Singh and R. M. Zingg, *Wolf Children and Feral Man* (New York, 1940); A. Gesell, *Wolf Child and Human Child* (New York, 1940).

2. Asked 'How often do wolves den in deserted ant mounds?' Singh replied that 'this was the only white-ant mound known to be lived in by the wolves'.

3. For a few years during World War II the Nationalist Socialists of Germany, since they behaved so inhumanly, were viewed as sub-human, and theories that all Nazis were insane found ready acceptance and were defended by psychiatrists.

4. While Pakenham-Walsh, who knew the area, spoke only of 'baby children', R. R. Gates, in his introductory remarks to the report on the wolf girls, mentions that in the jungle areas of India 'female children are still occasionally exposed'.

5. R. R. Gates, in a footnote to the description of the children's crawling, mentions that A. Hrdlicka collected 387 such cases, mostly of white children of civilised parents. This suggests that crawling is neither unique nor necessarily due to feral rearing.

6. For example, one of our autistic, non-speaking girls at one period kept her eyes shut tight for days. This in no way impaired her ability to find her way about, even when we tentatively put some obstacles in her normal course in the hope of inducing her to open her eyes. This she did not do but sensed exactly the place where the obstruction was put and circumnavigated it.

7. Mahler describes how one of her psychotic patients deliberately scorched her own lips with a cigarette lighter and showed no reaction, and she goes on to say that this child's sensitivity seemed grossly below normal and to explain this as an 'indication of the lack or deficiency of peripheral cathexis in autistic child patients'.

The development, maintenance and change of attitudes of an individual and the complex relationships between a person's attitudes and his behaviour may be considered a core problem of social psychology. This chapter complements the preceding sections – those on 'socialisation' and 'personality development' – since attitudes are learned through socialisation and may be considered part of an individual's personality.

Starting from the premise that we may in public voice opinions or attitudes without accepting them privately, H. C. Kelman discusses in his influential article on 'Three Processes of Social Influence' three distinct processes (which he calls 'compliance', 'identification' and 'internalisation'), each taking place under different kinds of conditions and leading to differen degrees of attitude change and persistence of the new attitude.

Bruno Bettelheim and Morris Janowitz, in 'Ethnic Tolerance: A Function of Social and Personal Control', found in their study of Chicago ex-servicemen after World War II that ethnic intolerance, that is intolerance towards people of a different race or cultural background, was most highly concentrated in those men who had been unsuccessful, while those who had risen showed more tolerance. The paper also discusses different types of controls (through societal pressures, rational self-control or guilt feelings) which may affect the expression of intolerance. Intolerance, then, can be accounted for by both sociological and psychological factors.

The paper by Irving Sarnoff, Daniel Katz and Charles McClintock on 'Attitude Change: Procedures and Motivating Patterns' shows that attitudes may fulfil different functions for the individual and may have different origins. This in turn has consequences for the strategies to be adopted in attempting to change a person's attitudes, and some experimental evidence testing the differential effectiveness of various methods of attitude change is discussed.

THREE PROCESSES OF SOCIAL INFLUENCE

H. C. Kelman

Social influence has been a central area of concern for experimental social psychology almost since its beginnings. Three general research traditions in this area can be distinguished: (1) the study of social influences on judgements, stemming from the earlier work on prestige suggestion; (2) the study of social influences arising from small-group interaction; and (3) the study of social influences arising from persuasive communications. In recent years, there has been a considerable convergence between these three traditions, going hand in hand with an increased interest in developing general principles of social influence and socially induced behaviour change.

One result of these developments has been that many investigators found it necessary to make qualitative distinctions between different types of influence. In some cases, these distinctions arose primarily out of the observation that social influence may have qualitatively different effects, that it may produce different kinds of change. For example, under some conditions it may result in mere public conformity – in superficial changes on a verbal or overt level without accompanying changes in belief; in other situations it may result in private acceptance – in a change that is more general, more durable, more integrated with the person's own values. Other investigators found it necessary to make distinctions because they observed that influence may occur for different reasons, that it may arise out of different motivations and orientations. For example, uuder some conditions influence may be primarily informational – the subject may conform to the influencing person or group because he views him as a source of valid information; in other situations influence may be primarily normative – the subject may conform in order to meet the positive expectations of the influencing person or group.

My own work can be viewed in the general context that I have outlined here. I started out with the distinction between public conformity and private acceptance, and tried to establish some of the distinct determinants of each. I became dissatisfied with this dichotomy as I began to look at important examples of social influence that could not be encompassed by it. I was especially impressed with the accounts of ideological conversion of the 'true believer' variety, and with the recent accounts of 'brainwashing',

particularly the Chinese Communist methods of 'thought reform'. It is apparent that these experiences do not simply involve public conformity, but that indeed they produce a change in underlying beliefs. But it is equally apparent that they do not produce what we would usually consider private acceptance – changes that are in some sense integrated with the person's own value system and that have become independent of the external source. Rather, they seem to produce new beliefs that are isolated from the rest of the person's values and that are highly dependent on external support.

These considerations eventually led me to distinguish three processes of social influence, each characterised by a distinct set of antecedent and a distinct set of consequent conditions. I have called these processes *compliance*, *identification* and *internalisation*.

THREE PROCESSES OF SOCIAL INFLUENCE

Compliance can be said to occur when an individual accepts influence from another person or from a group because he hopes to achieve a favourable reaction from the other. He may be interested in attaining certain specific rewards or in avoiding certain specific punishments that the influencing agent controls. For example, an individual may make a special effort to express only 'correct' opinions in order to gain admission into a particular group or social set, or in order to avoid being fired from his government job. Or, the individual may be concerned with gaining approval or avoiding disapproval from the influencing agent in a more general way. For example, some individuals may compulsively try to say the expected thing in all situations and please everyone with whom they come in contact, out of a disproportionate need for favourable responses from others of a direct and immediate kind. In any event, when the individual complies, he does what the agent wants him to do – or what he thinks the agent wants him to do – because he sees this as a way of achieving a desired response from him. He does not adopt the induced behaviour – for example, a particular opinion response – because he believes in its content, but because it is instrumental in the production of a satisfying social effect. What the individual learns, essentially, is to say or do the expected thing in special situations, regardless of what his private beliefs may be. Opinions adopted through compliance should be expressed only when the person's behaviour is observable by the influencing agent.

Identification can be said to occur when an individual adopts behaviour derived from another person or a group because this behaviour is associated with a satisfying self-defining relationship to this person or group. By a self-defining relationship I mean a role relationship that forms a part of the person's self-image. Accepting influence through identification, then, is a way of establishing or maintaining the desired relationship to the other, and the self-definition that is anchored in this relationship.

The relationship that an individual tries to establish or maintain through identification may take different forms. It may take the form of classical identification, that is, of a relationship in which the individual takes over all or part of the role of the influencing agent. To the extent to which such a relationship exists, the individual defines his own role in terms of the role of the other. He attempts to be like or actually to *be* the other person. By saying what the other says, doing what he does, believing what he believes, the individual maintains this relationship and the satisfying self-definition that it provides him. An influencing agent who is likely to be an attractive object for such a relationship is one who occupies a role desired by the individual – who possesses those characteristics that the individual himself lacks – such as control in a situation in which the individual is helpless, direction in a situation in which he is disoriented, or belongingness in a situation in which he is isolated.

The behaviour of the brainwashed prisoner in Communist China provides one example of this type of identification. By adopting the attitudes and beliefs of the prison authorities – including *their* evaluation of *him* – he attempts to retain his identity, which has been subjected to severe threats. But this kind of identification does not occur only in such severe crisis situations. It can also be observed, for example, in the context of socialisation of children, where the taking over of parental attitudes and actions is a normal, and probably essential, part of personality development. The more or less conscious efforts involved when an individual learns to play a desired occupational role and imitates an appropriate role model would also exemplify this process. Here, of course, the individual is much more selective in the attitudes and actions he takes over from the other person. What is at stake is not his basic

sense of identity or the stability of his self-concept, but rather his more limited 'professional identity'.

The self-defining relationship that an individual tries to establish or maintain through identification may also take the form of a reciprocal role relationship – that is, of a relationship in which the roles of the two parties are defined with reference to one another. An individual may be involved in a reciprocal relationship with another specific individual, as in a friendship relationship between two people. Or he may enact a social role which is defined with reference to another (reciprocal) role, as in the relationship between patient and doctor. A reciprocal role relationship can be maintained only if the participants have mutually shared expectations of one another's behaviour. Thus, if an individual finds a particular relationship satisfying, he will tend to behave in such a way as to meet the expectations of the other. In other words, he will tend to behave in line with the requirements of this particular relationship. This should be true regardless of whether the other is watching or not: quite apart from the reactions of the other, it is important to the individual's own self-concept to meet the expectations of his friendship role, for example, or those of his occupational role.

Thus, the acceptance of influence through identification should take place when the person sees the induced behaviour as relevant to and required by a reciprocal role relationship in which he is a participant. Acceptance of influence based on a reciprocal role relationship is similar to that involved in classical identification in that it is a way of establishing or maintaining a satisfying self-defining relationship to another. The nature of the relationship differs, of course. In one case it is a relationship of identity; in the other, one of reciprocity. In the case of reciprocal role relationships the individual is not identifying with the other in the sense of taking over *his* identity, but in the sense of emphatically reacting in terms of the other person's expectation, feelings or needs.

Identification may also serve to maintain an individual's relationship to a group in which his self-definition is anchored. Such a relationship may have elements of classical identification as well as of reciprocal roles: to maintain his self-definition as a group member, an individual, typically, has to model his behaviour along particular lines and has to meet the expectations of his fellow members. An example of identification with a group would be the member of the Communist Party who derives strength and a sense of identity from his self-definition as part of the vanguard of the proletarian revolution and as an agent of historical destiny. A similar process, but at a low degree of intensity, is probably involved in many of the conventions that people acquire as part of their socialisation into a particular group.

Identification is similar to compliance in that the individual does not adopt the induced behaviour because its content *per se* is intrinsically satisfying. Identification differs from compliance, however, in that the individual actually believes in the opinions and actions that he adopts. The behaviour is accepted both publicly and privately, and its manifestation does not depend on observability by the influencing agent. It does depend, however, on the role that an individual takes at any given moment in time. Only when the appropriate role is activated – only when the individual is acting within the relationship upon which the identification is based – will the induced opinions be expressed. The individual is not primarily concerned with pleasing the other, with giving him what he wants (as in compliance), but he is concerned with meeting the other's expectations for his own role performance. Thus, opinions adopted through identification do remain tied to the external source and dependent on social support. They are not integrated with the individual's value system, but rather tend to be isolated from the rest of his values – to remain encapsulated.

Finally, *internalisation* can be said to occur when an individual accepts influence because the induced behaviour is congruent with his value system. It is the content of the induced behaviour that is intrinsically rewarding here. The individual adopts it because he finds it useful for the solution of a problem, or because it is congenial to his own orientation, or because it is demanded by his own values – in short, because he perceives it as inherently conducive to the maximisation of his values. The characteristics of the influencing agent do play an important role in internalisation, but the crucial dimension here – as we shall see below – is the agent's credibility, that is, his relation to the content.

The most obvious examples of internalisation are those that involve the evaluation and acceptance of induced behaviour on rational grounds. A person may adopt the recommendations of an expert, for example, because he finds them relevant to his own

problems and congruent with his own values. Typically, when internalisation is involved, he will not accept these recommendations *in toto* but modify them to some degree so that they will fit his own unique situation. Or a visitor to a foreign country may be challenged by the different patterns of behaviour to which he is exposed, and he may decide to adopt them (again, selectively and in modified form) because he finds them more in keeping with his own values than the patterns in his home country. I am not implying, of course, that internalisation is always involved in the situations mentioned. One would speak of internalisation only if acceptance of influence took the particular form that I described.

Internalisation, however, does not necessarily involve the adoption of induced behaviour on rational grounds. I would not want to equate internalisation with rationality, even though the description of the process has decidedly rationalist overtones. For example, I would characterise as internalisation the adoption of beliefs because of their congruence with a value system that is basically *irrational*. Thus, an authoritarian individual may adopt certain racist attitudes because they fit into his paranoid, irrational view of the world. Presumably, what is involved here is internalisation, since it is the content of the induced behaviour and its relation to the person's value system that is satisfying. Similarly, it should be noted that congruence with a person's value system does not necessarily imply logical consistency. Behaviour would be congruent if, in some way or other, it fitted into the person's value system, if it seemed to belong there and be demanded by it.

It follows from this conception that behaviour adopted through internalisation is in some way – rational or otherwise – integrated with the individual's existing values. It becomes part of a personal system, as distinguished from a system of social-role expectations. Such behaviour gradually becomes independent of the external source. Its manifestation depends neither on observability by the influencing agent nor on the activation of the relevant role, but on the extent to which the underlying values have been made relevant by the issues under consideration. This does not mean that the individual will invariably express internalised opinions, regardless of the social situation. In any specific situation, he has to choose among competing values in the face of a variety of situational requirements. It does mean, however, that these opinions will at least enter into competition with other alternatives whenever they are relevant in content.

It should be stressed that the three processes are not mutually exclusive. While they have been defined in terms of pure cases, they do not generally occur in pure form in real-life situations. The examples that have been given are, at best, situations in which a particular process predominates and determines the central features of the interaction.

ANTECEDENTS AND CONSEQUENTS OF THE THREE PROCESSES

For each of the three processes, a distinct set of antecedents and a distinct set of consequents have been proposed. These are summarised in Table 8.1 below. First, with respect to the antecedents of the three processes, it should be noted that no systematic quantitative differences between them are hypothesised. The probability of each process is presented as a function of the same three determinants: the importance of the induction for the individual's goal achievement, the power of the influencing agent and the prepotency of the induced response. For each process, the magnitude of these determinants may vary over the entire range: each may be based on an induction with varying degrees of importance, on an influencing agent with varying degrees of power, and so on. The processes differ only in terms of the *qualitative* form that these determinants take. They differ, as can be seen in the table, in terms of the *basis* for the importance of the induction, the *source* of the influencing agent's power, and the *manner* of achieving prepotency of the induced response.

1. The processes can be distinguished in terms of the basis for the importance of the induction, that is, in terms of the nature of the motivational system that is activated in the influence situation. What is it about the influence situation that makes it important, that makes it relevant to the individual's goals? What are the primary concerns that the individual brings to the situation or that are aroused by it? The differences between the three processes in this respect are implicit in the descriptions of the processes given above: (*a*) To the extent that the individual is concerned – for whatever reason – with the *social effect* of his behaviour, influence will tend to take the form of compliance. (*b*) To the extent that he

is concerned with the *social anchorage* of his behaviour, influence will tend to take the form of identification. (*c*) To the extent that he is concerned with the *value congruence* of his behaviour (rational or otherwise), influence will tend to take the form of internalisation.

TABLE 8.1

SUMMARY OF THE DISTINCTION BETWEEN THE THREE PROCESSES

	Compliance	*Identification*	*Internalisation*
Antecedents:			
1. Basis for the *importance of the induction*	Concern with social effect of behaviour	Concern with social anchorage of behaviour	Concern with value congruence of behaviour
2. Source of *power of the influencing agent*	Means control	Attractiveness	Credibility
3. Manner of achieving *prepotency of the induced response*	Limitation of choice behaviour	Delineation of role requirements	Reorganisation of means–ends framework
Consequents:			
1. Conditions of performance of induced response	Surveillance by influencing agent	Salience of relationship to agent	Relevance of values to issue
2. Conditions of change and extinction of induced response	Changed perception of conditions for social rewards	Changed perception of conditions for satisfying self-defining relationships	Changed perception of conditions for value maximisation
3. Type of behaviour system in which induced response is embedded	External demands of a specific setting	Expectations defining a specific role	Person's value system

2. A difference between the three processes in terms of the source of the influencing agent's power is hypothesised. (*a*) To the extent that the agent's power is based on his *means control*, influence will tend to take the form of compliance. An agent possesses means control if he is in a position to supply or withhold means needed by the individual for the achievement of his goals. The perception of means control may depend on the agent's *actual* control over specific rewards and punishment, or on his *potential* control, which would be related to his position in the social structure (his status, authority or general prestige). (*b*) To the extent that the agent's power is based on his *attractiveness*, influence will tend to take the form of identification. An agent is attractive if he occupies a role which the individual himself desires or if he occupies a role reciprocal to one the individual wants to establish or maintain. The term 'attractiveness', as used here, does not refer to the possession of qualities that make a person likeable, but rather to the possession of qualities on the part of the agent that make a continued relationship to him particularly desirable. In other words, an agent is attractive when the individual is able to derive satisfaction from a self-definition with reference to him. (*c*) To the extent that the agent's power is based on his *credibility*, influence will tend to take the form of internalisation. An agent possesses credibility if his statements are considered truthful and valid, and hence worthy of serious consideration. Hovland, Janis and Kelley distinguish two bases for credibility: expertness and trustworthiness. In other words, an agent may be perceived as possessing credibility because he is likely to *know* the truth, or because he is likely to *tell* the truth. Trustworthiness, in turn, may be related to overall respect, like-mindedness and lack of vested interest.

3. It is proposed that the three processes differ in terms of the way in which prepotency is achieved. (*a*) To the extent that the induced response becomes prepotent – that is, becomes a 'distinguished path' relative to alternative response possibilities – because the individual's choice behaviour is limited, influence will tend to take the form of compliance. This may happen if the individual is pressured into the induced response, or if alternative responses are blocked. The induced response thus becomes prepotent because it is, essentially, the only response permitted: the individual sees himself as having no choice and as being restricted to this particular alternative. (*b*) To the extent that the induced response becomes prepotent because the requirements of a particular role are delineated, influence will tend to take the form of identification. This may happen if the situation is defined in terms of a particular role relationship and the demands of that role are more or less clearly specified; for instance, if this role is made

especially salient and the expectation deriving from it dominates the field. Or it may happen if alternative roles are made ineffective because the situation is ambiguous and consensual validation is lacking. The induced response thus becomes prepotent because it is one of the few alternatives available to the individual: his choice behaviour may be unrestricted, but his opportunity for selecting alternative responses is limited by the fact that he is operating exclusively from the point of view of a particular role system. (*c*) Finally, to the extent that the induced response becomes prepotent because there has been a reorganisation in the individual's conception of means–ends relationships, influence will tend to take the form of internalisation. This may happen if the implications of the induced response for certain important values – implications of which the individual had been unaware heretofore – are brought out, or if the advantages of the induced response as a path to the individual's goals, compared to the various alternatives that are available, are made apparent. The induced response thus becomes prepotent because it has taken on a new meaning: as the relationships between various means and ends become restructured, it emerges as the preferred course of action in terms of the person's own values.

Depending, then, on the nature of these three antecedents, the influence process will take the form of compliance, identification or internalisation. Each of these corresponds to a characteristic pattern of internal responses – thoughts and feelings – in which the individual engages as he accepts influence. The resulting changes will, in turn, be different for the three processes, as indicated in the second half of the table. Here, again, it is assumed that there are no systematic quantitative differences between the processes, but rather qualitative variations in the subsequent histories of behaviour adopted through each process.

1. It is proposed that the processes differ in terms of the subsequent conditions under which the induced response will be performed or expressed. (*a*) When an individual adopts an induced response through compliance, he tends to perform it only under conditions of *surveillance* by the influencing agent. These conditions are met if the agent is physically present, or if he is likely to find out about the individual's actions. (*b*) When an individual adopts an induced response through identification, he tends to perform it only under conditions of *salience* of his relationship to the agent. That is, the occurrence of the behaviour will depend on the extent to which the person's relationship to the agent has been engaged in the situation. Somehow this relationship has to be brought into focus and the individual has to be acting within the particular role that is involved in the identification. This does not necessarily mean, however, that he is consciously aware of the relationship; the role can be activated without such awareness. (*c*) When an individual adopts an induced response through internalisation, he tends to perform it under conditions of *relevance of the values* that were initially involved in the influence situation. The behaviour will tend to occur whenever these values are activated by the issues under consideration in a given situation, quite regardless of surveillance or salience of the influencing agent. This does not mean, of course, that the behaviour will occur every time it becomes relevant. It may be out-competed by other responses in certain situations. The probability of occurrence with a given degree of issue relevance will depend on the strength of the internalised behaviour.

2. It is hypothesised that responses adopted through the three processes will differ in terms of the conditions under which they will subsequently be abandoned or changed. (*a*) A response adopted through compliance will be abandoned if it is no longer perceived as the best path towards the attainment of social rewards. (*b*) A response adopted through identification will be abandoned if it is no longer perceived as the best path towards the maintenance or establishment of satisfying self-defining relationships. (*c*) A response adopted through internalisation will be abandoned if it is no longer perceived as the best path towards the maximisation of the individual's values.

3. Finally, it is hypothesised that responses adopted through the three processes will differ from each other along certain qualitative dimensions. These can best be summarised, perhaps, by referring to the type of behaviour system in which the induced response is embedded. (*a*) Behaviour adopted through compliance is part of a system of external demands that characterise a specific setting. In other words, it is part of the rules of conduct that an individual learns in order to get along in a particular situation or series of situations. The behaviour tends to be related to the person's values only in an instrumental rather than an intrinsic way. As long as opinions, for example, remain at that level, the individual will tend to regard them as not really representative of his true beliefs.

(*b*) Behaviour adopted through identification is part of a system of expectations defining a particular role – whether this is the role of the other which he is taking over, or a role reciprocal to the other's. This behaviour will be regarded by the person as representing himself, and may in fact form an important aspect of himself. It will tend to be isolated, however, from the rest of the person's values – to have little interplay with them. In extreme cases, the system in which the induced response is embedded may be encapsulated and function almost like a foreign body within the person. The induced responses here will be relatively inflexible and stereotyped. (*c*) Behaviour adopted through internalisation is part of an internal system. It is fitted into the person's basic framework of values and is congruent with it. This does not imply complete consistency: the degree of consistency can vary for different individuals and different areas of behaviour. It does mean, however, that there is some interplay between the new beliefs and the rest of the person's values. The new behaviour can serve to modify existing beliefs and can in turn be modified by them. As a result of this interaction, behaviour adopted through internalisation will tend to be relatively idiosyncratic, flexible, complex and differentiated.

ETHNIC TOLERANCE: A FUNCTION OF SOCIAL AND PERSONAL CONTROL

Bruno Bettelheim and Morris Janowitz

IN this study of ethnic intolerance we attempt to throw light on the principles of group hostility in general and on ethnic hostility as a special subtype.

The four main hypotheses that the research sought to test were based on sociological theory and dynamic psychology. They were: (1) hostility towards out-groups is a function of the hostile individual's feeling that he has suffered deprivations in the past; (2) such hostility towards out-groups is a function of the hostile individual's anxiety in anticipation of future tasks; (3) the individual blames out-groups for his failure at mastery and projects undesirable characteristics denied in himself upon members of the out-group because of inadequate personal and social controls which favour irrational discharge and evasion rather than rational action; (4) ethnic intolerance can be viewed in terms of the individual's position within the social structure either statically or dynamically. It was assumed that ethnic intolerance was related more to the individual's dynamic movement within the structure of society than to his position at a particular moment. No claim is made that these hypotheses are universally applicable, but they seemed useful in understanding hostility in modern industrialised communities.

A major premise of the study was that persons who believe they have undergone deprivations are disposed to ethnic intolerance. It seemed plausible to study ex-soldiers, since they had suffered deprivations in varying degrees and might be especially responsive to the appeal of intolerance. A random sample of 150 male war veterans, all residents of Chicago, was studied. Former officers were eliminated from the study, since their experiences were at variance with those of enlisted men and since most of them came from social and economic backgrounds which differed from those of enlisted men. Hence the sample tended more adequately to represent the economic lower and lower-middle classes. Members of those major ethnic groups towards which hostility is projected were not included – that is, Negroes, Jews, Chinese, Japanese and Mexicans.

The data were obtained through intensive interviews in which free associations were always encouraged. The interviewers were psychiatrically trained social workers, experienced in public-opinion surveying. The wide range of personal data sought and the special problems of building rapport before gathering data on ethnic attitudes required long interviews which took from four to seven hours and in several cases were carried on in two sessions. The veterans were offered ample opportunity to express personal views on many issues and to recount their war-time experiences before ethnic minorities were mentioned.

On the basis of an exploratory study we found it necessary to distinguish four types of veterans with respect to their ethnic attitudes. For the sake of brevity, only the four types of anti-Semite are mentioned, but a parallel classification as regards anti-Negro attitudes was also developed. These four types of anti-Semite were designated as *intensely anti-Semitic, outspoken anti-*

Semitic, stereotyped anti-Semitic and *tolerant* towards Jews and were characterised as follows: (1) The *intensely anti-Semitic* veteran was spontaneously outspoken in expressing a preference for restrictive action against the Jews even before the subject was raised. (2) The *outspoken anti-Semitic* man revealed no spontaneous preference for restrictive action against the Jews. Instead, outspoken hostility towards the Jews emerged only towards the end of the interview when he was directly questioned. As in the case of the intensely anti-Semitic veteran, his thinking contained a wide range of unfavourable stereotypes. (3) The *stereotyped anti-Semitic* man expressed no preference for hostile or restrictive action against the Jews even when questioned directly. Instead, he merely expressed a variety of stereotyped notions about the Jews, including some which were not necessarily unfavourable from his point of view. (4) The *tolerant* veteran revealed no elaborately stereotyped beliefs about the Jews (among the statements of even the most tolerant veterans isolated stereotypes might from time to time be found). Moreover, not even when questioned directly did he advocate restrictive action against the Jews.

The interview situation was so constructed that the responses to questions would permit a clear discrimination between these four types of ethnic intolerance. The first portion of the interview was designed to offer the men an opportunity for spontaneous expression of hostility against minorities without bringing this subject to their attention. In a second portion, especially in connection with Army experiences, ample opportunity was offered to display stereotyped thinking by asking, for example, who the 'goldbrickers' or troublemakers had been. Only the last portion contained direct questions on ethnic minorities. There the stimuli 'Negro' and 'Jew' were introduced to determine which men were consistently tolerant. First it was asked what kinds of soldiers they made, next what the subject thought of social and economic association with them and then what his views were on possible changes in the current patterns of inter-ethnic relations. Table 8.2 shows the distribution of degrees of intolerance.

We tried to determine whether the men's social and economic history could account for their ethnic intolerance. Among the characteristics studied were age, education, religion, political affiliation, income and social status. But the data indicate that – subject to certain limitations – these factors of themselves do not seem to account for differences in the degree or nature of intolerance.

TABLE 8.2
DISTRIBUTION OF INTOLERANCE

	Anti-Semitic		*Anti-Negro*	
	No.	*Per cent*	*No.*	*Per cent*
Tolerant	61	41	12	8
Stereotyped	42	28	40	27
Outspoken	41	27	74	49
Intense	6	4	24	16
Total	150	100	150	100

Table 8.3, for example, shows that no statistically significant relation exists between income and socio-economic status on the one hand, and intensity of anti-Semitism on the other.[1] The same was true for such other categories as education, age and religious affiliation. Which newspaper, magazine or radio programme the men favoured was also unrelated to the intensity of ethnic hostility. The pattern of anti-Negro distribution was similar.

SOCIAL MOBILITY

The picture changes, however, if a static concept of social status is replaced by the dynamic concept of social mobility. It was possible to gather precise data on the social mobility of 130 veterans. They were rated as having experienced downward mobility or upward mobility if they had moved at least one grade up or down on the Alba Edward's socio-economic scale when compared with their previous civilian employment.

Table 8.4 shows that ethnic hostility was most highly concentrated in the downwardly mobile group, while the pattern was *significantly* reversed for those who had risen in their social position. Those who had experienced no change presented a picture somewhat in the middle; the relationship between ethnic intolerance and social mobility (as defined in this study) was also present when educational level was held constant.

The group which was static showed the highest concentration of stereotyped opinions – that is, they were 'middle-of-the-roaders' with regard to anti-Semitism. Over 70 per cent of the stereotyped anti-Semites were found in this middle category. This illuminates

the relation between mobility and intolerance. On the other hand, the no-mobility group was most generally in the outspokenly anti-Negro category. This supplies another crude index of the limits of intolerance towards minority groups in a northern urban industrial community. In the case of the Jew the social norms were most likely to produce merely stereotyped thinking, while it was correspondingly 'normal' to be outspoken in one's hostility towards the Negro.

Table 8.3

Correlates of Anti-Semitism in Per Cent

Total cases	*Tolerant* (61)	*Stereotyped* (42)	*Outspoken and intense* (47)	*Total* (150)	*No.*
Age:					
Under 28	44	27	29	100	94
29–36	34	30	36	100	56
Education:					
Did not complete high school	35	31	34	100	65
Completed high school	39	28	33	100	46
Some college or more	51	23	26	100	39
Religion:*					
Catholic	40	28	32	100	103
Protestant	48	25	27	100	33
No present religious denomination	33⅓	33⅓	33⅓	100	12
Current salary:					
Up to $2500	39	33	28	100	59
$2500 to $3000	39	24	37	100	43
$3000 and over	43	18	39	100	28
Not applicable	45	35	20	100	20
Socio-economic status:					
Top four groups	42	24	34	100	70
Semi-skilled and unskilled	38	33	29	100	80

* Two cases of Greek Orthodox not included.

In view of the association between upward social mobility and tolerance, the few cases (14) who displayed both upward mobility and were outspokenly anti-Semitic warrant special attention. The actual income gains associated with upward mobility reveal that the men who were both outspokenly anti-Semitic and upwardly mobile tended to be considerably more mobile than the others. This may be tentatively explained by the fact that sharp upward mobility is likely to be associated with marked aggressiveness in general. The data, particularly on those in the group downwardly mobile, suggest that to understand intolerance it is less important to concentrate on the social and economic background of the individual than to investigate the character of his social mobility.

Feeling of Deprivation

Whatever their social and economic life-histories had been, all the men interviewed had one common experience – the Army. Reactions to comparable war-time deprivations thus afforded a unique opportunity to examine the hypothesis that the individual who suffers deprivation tries to restore his integration and self-control by the expression of hostility, one form of which may be ethnic hostility. But here a sharp distinction must be introduced between *actual* deprivations experienced and his *feelings* of deprivation. Whether the men reacted favourably to Army life primarily because they experienced relief from the insecurities of civilian life was also pertinent.

Army experiences which involved *objective* deprivations were found not related to differential degrees of ethnic intolerance (combat versus non-combat service, wounds, length of service, etc.). On the other hand, a clear association emerged between the display of *feelings* of deprivation and outspoken or intense anti-Semitic and anti-Negro attitudes.

who declared themselves particularly attached to Army life displayed a high concentration of intolerance.

The judging of one's war experiences as depriving or not is a function of the individual's total personality and of the adequacy of his adjustive mechanisms. The interview records of those who seemed gratified by Army life revealed that they were also the

TABLE 8.4
INTOLERANCE AND MOBILITY

	Downward mobility		*No mobility*		*Upward mobility*		*Total*	
	No.	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*
Anti-Semitic:								
Tolerant	2	11	25	37	22	50	49	38
Stereotyped	3	17	26	38	8	18	37	28
Outspoken and intense	13	72	17	25	14	32	44	34
Anti-Negro:								
Tolerant and stereotyped	5	28	18	26	22	50	45	34
Outspoken	5	28	40	59	17	39	62	48
Intense	8	44	10	15	5	11	23	18
Total	18	100	68	100	44	100	130	100

On the basis of a content analysis it was found that it was possible to make reliable decisions as to whether the veterans (1) accepted it in a matter-of-fact way, (2) were embittered about Army life, or (3) were attached to it or gratified by it. The overwhelming majority of those who were tolerant, regardless of the specific content of their war-time experience, had an attitude of acceptance towards Army life, while the intolerant veteran presented a completely reversed picture (see Table 8.5). The latter were overwhelmingly embittered by Army life. In addition, those

men who described themselves as economically and socially deprived before induction; they seem to have been poorly adjusted to civilian society and to have found gratification and release in the particular adventure and comradeship of Army life.

CONTROLS FOR TOLERANCE

There seems little doubt that frustrating social experiences and the inability to integrate them account to a large degree for those

TABLE 8.5
ACCEPTANCE OF ARMY

	Tolerant		*Stereotyped*		*Outspoken and intense*		*Total*	
	No.	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*
Accepted Army life	44	81	21	64	6	17	71	50
Embittered towards Army life	6	11	7	21	20	56	33	35
Attached to or gratified by Army life	4	8	5	15	10	27	19	15
Total	54	100	33	100	36	100	123	100

aggressions which are vented in ethnic hostility. While our investigation could not ascertain which particular experiences accounted for the men's frustration, it permitted us to ascertain their readiness to submit in general to the existing controls by society. If, by and large, they accepted social institutions, it seems reasonable to assume that such acceptance implied a willingness to control their own aggressive tendencies for the sake of society. Or, oversimplifying complex emotional tendencies, one might say that those men who felt that society fulfilled its task in protecting them against unavoidable frustrations were also those who, in return, were willing to come to terms with society by controlling their aggressive tendencies as society demands. Hence, the hypothesis correlating the men's acceptance or rejection of society with their ethnic attitudes had to be tested. The Army is only one of many social institutions. The postulated association between intolerance and the rejection of social controls, which was central in terms of this study, had to be investigated for a number of other institutions as well.

Control, technically speaking, is the ability to store tension internally or to discharge it in socially constructive action rather than in unwarranted hostile action. The predominant mechanisms of control which a person uses for dealing with inner tensions are among the most important elements characterising his personality. Each of these mechanisms of control is more or less adequate for containing a particular type of aggression generated in the individual by anxiety. These controls or restraints remain adequate only if the level of tension does not become overpowering, thereby creating unmasterable anxiety. It will not suffice to investigate the association between control and tolerance in general; it is necessary to discriminate between tolerance as it relates to three types of control over hostile tendencies: (1) external or social control, (2) superego or conscience control, and (3) rational self-control or ego control.

Religion may serve as the prototype of an institution, the acceptance of, or submission to, which was found to be related to tolerance. Unquestioning acceptance of religious values indicates that the individual tends to rely on a type of control in which he is guided by traditional and non-rational external social forces. In contrast, control is exercised not by the minister or the priest but originates within the person, although such inner control may have come initially from their teachings. If the moral teachings of the church are accepted by the individual not through fear of damnation or of societal disapproval but because he considers them absolute standards of behaviour independent of external threats of approval, then we say that the individual has 'internalised' these moral precepts. They have become an internal control, but a control which is still only partially conscious and only partly rational. Such control is exercised over the individual by his 'conscience', or, technically speaking, by his superego.

Markedly different from *external* control through outside institutions and from *superego* control, which also depends for its effectiveness on props in the external world (such as parental images or institutionalised religion), is the rational control of irrational tendencies which forces them into consciousness and then deals with them along purely rational lines. The latter may be termed 'ego control'. In actuality, the three types of control are nearly always coexistent, and in each individual case control will depend in varying degrees on all three – external, superego and ego control. In the men studied, wherever control was present it was overwhelmingly the result of a combination of external and superego control, with the first being dominant. Only a few men were also motivated by ego control, and in even fewer was ego control dominant over superego or external control. Hence a study of external, i.e. societal, control was the only one which promised to permit insight into the correlation between acceptance of, or submission to, social control and ethnic intolerance for this particular group.

The analysis of religious attitudes indicated that veterans who had stable religious convictions tended to be the more tolerant. When the political party system was viewed as another norm-setting institution, a similar relationship of at least partial acceptance or consensus with this basic institution was found to be associated with tolerance. Whether the veteran was Democratic or Republican was in no way indicative of his attitude towards minorities. But the veteran who rejected or condemned both parties ('they are both crooks') tended to be the most hostile towards minorities.

Thus not only greater stability in societal status but the very existence of stable religious and political affiliations as well proved to be correlated with tolerance. These phenomena are indicative

of the tolerant individual's relatively greater control over his instinctual tendencies, controls which are strong enough to prevent immediate discharge of tension in asocial action. Such delay in the discharge of tension permits its canalisation into socially more acceptable outlets.

To explore more fully this relationship between tolerance and control, the responses to other symbols of societal authority which signify *external* control of the individual were also investigated. Two groups of institutions were analysed separately. The first group, that of Army control through discipline and officers' authority, is discussed below. The second group was composed of significant representatives of civilian authority to which the men were relatively subject at the time of the interview.

Four institutions were singled out as being most relevant. They were: (1) the administration of veterans' affairs; (2) the political party system; (3) the federal government; and (4) the economic system, as defined by the subjects themselves.

The veterans' views of each of these institutions were quite complex and in some respects ambivalent. Nevertheless, it was possible to analyse attitudes towards them on a continuum of acceptance, rejection or intermediate.

When acceptance or rejection of the four representative institutions was compared with the degree of anti-Semitism (Table 8.6), it appeared that only an insignificant percentage of the tolerant men rejected them, while nearly half the outspoken and intense anti-Semites did so. This is in marked contrast, for example, to studies of certain types of college students, in whom radical rejection of authority is combined with liberalism towards minority groups.

Controls, it may be said, are not internalised by merely accepting society. On the contrary, general attitudes of accepting existing society and its institutions are the result of previous internalisation of societal values as personally transmitted by parents, teachers and peers. Hence the acceptance of individuals who are representatives of societal values should have been more closely related to internal control than the acceptance of discipline in general, which is more characteristic of external control. Attitudes towards officers seemed suitable gauges for the individual's attitudes towards control. Incidentally, most of the men evaluated their officers on the basis of personal quality, their moral authority, and not on the basis of their punitive power.

The tolerant veteran appeared able to maintain better relations with his officers; he was more willing to accept the authority and discipline of the Army as represented by them. In general, his attitude was reasonable. When queried as to how the fellows in their outfits got along with the officers, tolerant veterans were significantly more prone to claim they got along well than were the intolerant men.

In the case of the Negro (Table 8.7), societal controls exercise a restraining influence only on what would be classified as violent, or 'intense', intolerance. Violence is generally disapproved of by the controlling institutions, while they approve, if not enforce, stereotyped and outspoken attitudes. The men who were strongly influenced by external controls were, in the majority, stereotyped and outspoken but not intense in their intolerance towards Negroes, as the present data show.

The division between those who rejected and those who accepted external control came between outspoken and intense attitudes

TABLE 8.6

ATTITUDES TOWARDS THE JEW AND TOWARDS CONTROLLING INSTITUTIONS

Attitude	*Tolerant*		*Stereotyped*		*Outspoken and intense*		*Total*	
	No.	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*
Accept	41	67	20	48	11	23	72	48
Intermediate	15	25	17	40	13	28	45	30
Reject	5	8	5	12	23	49	33	22
Total	61	100	42	100	47	100	150	100

towards Negroes. To score 'high' on the index of rejection for the four controlling institutions meant that an individual was likely to fall in the intensely anti-Negro category. Thus acceptance of external controls not only was inadequate in conditioning men to be tolerant of the Negroes but was not even enough to prevent them from holding outspoken views in that regard. It served only to restrain demands for violence.

They were more able to evaluate correctly the individuals whom they met, but they clung to stereotyped thinking about the rest of the discriminated group. In this way it remained possible to retain the stereotyped attitudes which permitted discharge of hostility despite actual experiences to the contrary. Such a limited amount of reality testing did not seem to be available to strongly biased individuals.

TABLE 8.7

ATTITUDES TOWARDS THE NEGRO AND TOWARDS CONTROLLING INSTITUTIONS

Attitude	*Tolerant*		*Stereotyped*		*Outspoken*		*Intense*		*Total*	
	No.	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*	*No.*	*Per cent*
Acceptance	9	75	19	48	38	51	6	25	72	48
Intermediate	2	17	16	40	23	31	4	17	45	30
Rejection	1	8	5	12	13	18	14	58	33	22
Total	12	100	40	100	74	100	24	100	150	100

STEREOTYPED THINKING

Precisely because most of the men in the sample based their restraint of aggressive tendencies on societal controls rather than on inner integration, some aggression remained uncontrolled. This the men needed to explain to themselves – and to others. For an explanation they fell back again on what society, or rather their associates, provided in the way of a justification for minority aggression. It has already been mentioned that most of the men voiced their ethnic attitudes in terms of stereotypes. The use of these stereotypes reveals a further influence – if not control – by society on ethnic attitudes and should therefore at least be mentioned.

One of the hypotheses of this study is that intolerance is a function of anxiety, frustration and deprivation, while the intolerant person's accusations are ways to justify his aggression. While the rationalisations for this intolerance must permit a minimum of reality testing, they will also condition the ways in which hostile feelings are discharged.

All intolerant veterans avoided reality testing to some degree, and each of them made statements about minorities which showed that he neglected the individual's uniquely personal characteristics – in short, he used stereotypes. As was to be expected, those who were moderately biased retained more ability to test reality.

Because the intolerant person's rationalisations are closely, although not obviously, connected with his reasons for intolerance, he must take care to protect them. On the other hand, they also reveal the nature of the anxieties which underlie them.

An examination of the five most frequent Negro and five most frequent Jewish stereotypes reveals strikingly different results, each set of which presents a more or less integrated pattern (see Tables 8.8 and 8.9). The composite pattern of stereotypes about Jews does not stress personally 'obnoxious' characteristics. In the main, they are represented in terms of a powerful, well-organised group which, by inference, threatens the subject.

On the other hand the stereotypes about the Negro stress the individual, personally 'offensive' characteristics of the Negro. As the stereotypes of the group characteristics of Jews implied a threat to the values and well-being of the intolerant white, so, too, those about the Negro were used to describe a conception of the Negro as a threat, particularly because the Negro was 'forcing out the whites'.

A comparison of the distribution of stereotypes applied to Jews and Negroes, as indicated by this enumeration, with those used by the National Socialists in Germany permits certain observations. In Germany the whole of the stereotypes, which in the United States were divided between Jews and Negroes, were applied to

the Jews. Thus in the United States, where two or more ethnic minorities are available, a tendency emerges to separate the stereotypes into two sets and to assign each of them to one minority group. One of these two sets indicates feelings of being anxious because of one minority's (the Jews') assumed power of overwhelming control. The other set of stereotypes shows feelings of

TABLE 8.8
STEREOTYPES CHARACTERISING JEWS

Stereotype	*No. of veterans mentioning stereotypes*
They are clannish; they help one another	37
They have the money	26
They control everything (or have an urge to control everything); they are running the country	24
They use underhanded or sharp business methods	24
They do not work; they do not do manual labour	19

TABLE 8.9
STEREOTYPES CHARACTERISING NEGROES

Stereotype	*No. of veterans mentioning stereotypes*
They are sloppy, dirty, filthy	53
They depreciate property	33
They are taking over; they are forcing out the whites	25
They are lazy; they are slackers in work	22
They are ignorant; have low intelligence	18
They have low character; they are immoral and dishonest	18

anxiety because of the second minority's (the Negroes') assumed ability to permit itself the enjoyment of primitive, socially unacceptable forms of gratification. Thus, of two minority groups which differ in physical characteristics, such as skin colour, the minority showing greater physical difference is used for projecting anxieties associated with dirtiness and sex desires. Conversely, the minority whose physical characteristics are more similar to those of the majority become a symbol for anxieties concerning overpowering control. If we apply the frame of reference of dynamic psychology to these observations, then these stereotypes permit further emphasis on the relation between tolerance and control. The individual who has achieved an integration or an inner balance between superego demands and instinctual, asocial strivings does not need to externalise either of them in a vain effort to establish a control that he does not possess. The intolerant man who cannot control his superego demands or instinctual drives projects them upon ethnic minorities as if, by fighting them in this way or by at least discharging excessive tension, he seeks to regain control over unconscious tendencies.

Actual experiences later in life, once the personality has been formed, seem relatively incapable of breaking down this delusional mechanism. Questioning revealed, for example, that although Army experience threw the men into new and varied contacts with Jews and frequently with Negroes, the stereotypes applied to the service of Jews and Negroes in the Army proved largely an extension of the conceptions of civilian life into Army experiences.

It seems reasonable to assume that, as long as anxiety and insecurity persist as a root of intolerance, the effort to dispel stereotypes by rational propaganda is at best a half-measure. On an individual level only greater personal integration combined with social and economic security seems to offer hope for better interethnic relations. Moreover, those who accept social controls are the more tolerant men, while they are also, relatively speaking, less tolerant of the Negro because Negro discrimination is more obviously condoned, both publicly and privately. This should lead, among other things, to additional efforts to change social practice in ways that will tangibly demonstrate that ethnic discrimination is contrary to the mores of society, a conviction which was very weak even among the more tolerant men.

Note

1. Where a significant difference is reported, it is at least at the 0·01 confidence limit.

ATTITUDE-CHANGE PROCEDURES AND MOTIVATING PATTERNS

Irving Sarnoff, Daniel Katz and Charles McClintock

A MAJOR difficulty in the field of attitude research has been the oversimplification of problems in terms of a narrow theory of motivation. The perceptual model of the Gestalt school sees the situation as one in which people strive towards a more inclusive and stable organisation of the psychological field; thus the individual tries to reconcile conflicting impressions, seeks to know what the world is like, and to make sense of it. According to this model, individuals will change their attitudes in rational fashion if they are presented with facts and reasons which accord with their own beliefs and assumptions. Another motivational model had followed the reward–punishment pattern and has seen attitudes as part of an adaptive response to the social world. Group norms become of first importance since the individual must have acceptance and support from his group. Finally, the personality theorist and clinician emphasise the internal dynamics underlying attitudes in which the individual's need to preserve his self-image and self-integrity become more important than external reward and punishment.

THE RATIONAL MODEL: THE SEARCH FOR MEANING

This approach emphasises the fact that people seek to understand their world and to achieve a consistent picture of its complexities. They attempt to deal with the realities of the matter which are within the framework of their own experiences and to deal with these realities in an understanding fashion. The blind emotionalism and the stupid irrationalities of a person are not necessarily as blind or as irrational if we took the trouble to understand his particular set of experiences and his frame of reference. 'Emotional processes are as a rule under the direction of cognitive factors, and are controlled by a trend to find relevant relations.'[1]

It is assumed that every person has a need to explore, to know about the external world. This need may be observed in the meanderings of naïve children, as well as in the controlled experiments of trained scientists. Often this quest for information seems to be a curiosity which is entirely self-contained. At other times, it appears to be in the service of other motives, such as survival or the desire to master the environment.

Considered in this light, attitudes may sometimes be a function of the range of information which has been accessible to the individual in regard to certain target objects. Let us assume, for example, that a curious child has heard of the word *Negro* for the first time. He goes to his parents for information and is told, among other things, that Negroes are bad people, stupid, dirty, animal-like. Assuming that he has heard no information to the contrary, that he has had no actual experience with Negroes, and that his parents have always supplied him with reliable facts about things, it follows that he will be inclined to accept this statement about Negroes. Even if the child is entirely free of other motivational needs to maintain this attitude, it is likely to persist until his contact with Negroes indicates to him that not all of them are stupid or until he is confronted by contrary information which he considers to be more reliable than that which his parents gave him.

REWARD AND PUNISHMENT

Some of the principles of learning theory may be applied to certain types of attitude formation. Specifically, these principles are involved when attitudes are adopted as a consequence of externally applied rewards and punishments. A white youngster in rural Mississippi may begin life on very favourable terms with his Negro playmates. Still, as he matures, his adjustment to and acceptance by the white community depend in some measure upon the extent to which he has come to share the prevalent anti-Negro biases. If he rejected these biases, he might be vulnerable to social ostracism and even more severe manifestations of disapproval.

EGO DEFENCE

In his social relationships every individual attempts to gratify his basic physiological needs and the socially acquired ones. Some individuals, because of the strength, maturity and integration of their egos, are able to perceive and tolerate the existence of their inner drives, including those which may be socially taboo. Such a person can, in terms of a realistic assessment of the social situation,

decide which impulses to express and which ones to suppress. It is possible for him to approach others in a straightforward manner; to react to them in terms of how they, as unique individuals, actually fulfil or frustrate his needs. His social attitudes are likely to be appropriate since he is able to differentiate clearly between what is going on inside him and what is emanating from others.

At the other end of the scale of ego strength there are persons who are obliged to use devious means of gratifying their impulses. In an attempt to resolve inner conflict, they minimise or obliterate certain aspects of their emotional life. This deflection of impulse from conscious awareness does not destroy it. On the contrary, it continues to press for overt expression and requires the individual to expend considerable energy in keeping it below the threshold of consciousness. Despite the most strenuous counter-measures (repression and other mechanisms of defence), the impulse tends to attain at least partial expression and gratification. This expression is achieved by means of symptom formation, a device which, however incapacitating it may be, serves a dual and somewhat paradoxical function:

1. It permits expression of the unconscious impulse.
2. It prevents the individual from becoming aware of the existence of the unconscious impulse.

Attitudes may thus function as ego defences, and may be viewed as symptoms. For example, the type of bigot described in the California studies is one who cannot accept his own hostile impulses.[2] By projecting them on to others, he gains gratification of his impulses while maintaining the fiction that these impulses originate in others rather than in himself.

Rather than attempting to elaborate any one of these three models to take account of all psychological phenomena concerned with attitudes, it seemed wiser to us to accept their essential contributions and to specify the conditions under which one or another of the theories best accounts for the phenomena under investigation. We feel that all people develop and alter their attitudes in terms of all three motivational contexts. Certain types of people, however, will be more likely to acquire attitudes on the basis of one kind of motivational pattern than will other people. Moreover, attitudes formed in the service of a given type of motivation will tend to show different characteristics than attitudes acquired in the interests of other motivational sources. In general, to take one important example, the dimension of the appropriateness or inappropriateness of an attitude, i.e. the amount of gross distortion of the target or referent of the attitude, will be related to the third type of motivation, ego defensiveness. Since the attitude has the function of protecting the individual from facing his internal conflicts, it will not be referred to the external world for reality testing. Moreover, any prediction of change or any attempts to modify attitudes must take account of these different motivational bases: (1) attitudes may be acquired in the interest of rationally structuring the individual's world and of testing what the world is like; (2) they may be formed as an adaptation to rewards and punishments imposed by the social situation; and (3) the attitude may be a function of the ego-defensive needs of the individual. It follows that the processes which produce a change in the first type of attitudinal structure will not necessarily be operative for the second or third types of structure. Where the individual's attitude is based upon his search for meaning, his attitude can be changed by giving him more information about the cognitive object. Where the attitude is based, however, upon a motive like group approval or desire to advance in the group, additional information which opposed this attitude may have no impact. Where the attitude reflects the individual's manner of resolving his inner conflicts, he will resist changes which would be in his own self-interest.

It should be noted, moreover, that attempts to change the ego-defensive individual which are based upon the logic of the first two approaches may actually reinforce his old attitudes. These appeals really assume either (*a*) the individual is interested in a more accurate, and more complete knowledge of the world, or (*b*) he is primarily concerned with maximising the satisfaction of his conscious needs. If, however, he is primarily concerned with avoiding a direct facing of his own internal conflicts, then he will protect himself from such a possibility the moment he senses an attempt to make him change. The well-known phenomenon of resistance is an example in point. The resistance generated to protect the ego results in a blanket rejection of the change situation, a perceptual distortion of what it is like and an emotional reinforcement of the attitude. Though we recognise the importance of resistance in other contexts, we have slighted its role in attitude

change. Hence, persuasion and propaganda can have negative rather than positive effects.

It is not our thesis, however, that ego defences cannot be breached, but that different procedures are involved in dealing with these defences than with non-conflicted motives.

ADAPTING CHANGE-PROCEDURES TO THE MOTIVATIONAL BASIS OF ATTITUDE

I. *Changing attitudes through attacking the cognitive object and the frame of reference in which it is perceived: the rational approach*

Inappropriate attitudes, though more characteristically acquired as ego defence, can also be acquired (*a*) in the interests of social rewards and punishments and (*b*) in the functioning of the perceptual-cognitive process itself. In the last-mentioned instance, the assumption is that the individual acquires beliefs either out of intellectual curiosity, reality-testing, or a desire for cognitive structure in terms of having to know what his world is really like. The belief or attitude may be inappropriate because the individual comes into contact with only a limited aspect of reality. In practice, of course, a number of these processes occur simultaneously. The individual takes on certain attitudes because he wants to be accepted as a group member, but the beliefs he acquires must make sense to him with respect to being accurate pictures of the world he knows.

The theory would be that when new facts are presented they result in cognitive reorganisation rather than exclusion or blocking. In addition, sophisticated cognitive theory assumes that perception occurs in a frame of reference which is the momentary product of a number of possible forces. Pressing internal needs or the objective nature of the situation may result in a temporary frame in which things are perceived in relation to that frame. Since people already possess definite standards for perceiving Negroes,[3] it is important, if we are to change their perceptions, to be sure that the dominant frame of reference at the time the materials are introduced is not the product of these old standards. Essentially, this calls for the activation of a 'new' frame without any reference to the old standards. Once the 'new' frame is operative, material bearing upon the old cognitive object can be introduced without the old standard operating as a censor. Specifically this can be accomplished through experimentation in the form of the creation of a frame of cultural relativity and cultural causation, and then the introduction of information about Negroes.

II. *Changing attitudes through the application of social rewards and punishments*

1. *The use of group norms.* One of the major reasons for attitude change is the desire to gain the social approval of others. It is natural that experiments in social psychology have concentrated heavily upon the manipulation of group norms. The desire for social approval is a broad term, however, covering both deep-lying affiliation needs and the immediate rewards of adapting to the group situation. We are concerned here primarily with this latter type of conformity, namely the acceptance of group standards because of their immediate reward character. For example, the individual who accepts the values of the group has increased his chances of being liked by his fellows, of moving up in the group structure or of sharing in the returns from group effort. In these terms, it would first be necessary to find out what the individual's relation to the group is and what types of satisfaction he derives from group membership. If we know these facts, then we are in a better position to predict changes in attitude as a result of a changed perception of group norms. Similarly we could predict those individuals who would tend to maintain their attitudinal position independent of the group. The assumption here is at the simple level that the more the rewards and the greater the sanctions the group has, the more effective will a changed perception of the group position be. In the persuasion pattern the attempt is made to achieve cognitive restructuring through making the individual see new possibilities in achieving his goals. In the use of group norms, the change is more external in that the group norm is changed and the individual merely moves with it.

2. *Restructuring the individual's value systems.* The great bulk of the efforts to change attitudes in the world of affairs is through persuasion and argumentation, in which an appeal is made to existing value structures. Some value system, other than that to which the target object is ordinarily connected, is invoked, and the necessity of meeting its requirements is stressed. In inducing change in the manner described under the *rational* approach, the emphasis was placed upon the perceptual side of the process, in

getting people to look at the old object in a new frame of reference. Thus subjects would have had to take the further step of restructuring their attitudes themselves. No direct attempt would be made to suggest that they should not pursue a discriminatory policy towards Negroes. In this second type of experiment, however, the attempt will be to restructure their attitudes for them through making the specific connections between the target object and a different set of values. This can be done (1) through exploiting a very powerful system, (2) through utilising more different values for the change than existed for the old attitudinal structure, and (3) through showing how the values tied to the target are blocked by social reality and are not in fact being achieved. In practice, all three methods are combined to produce change. From the point of view of content, the persuasive technique frequently relies upon demonstration of a fusion between the individual's self-interest and socially desirable goals. It further shows the feasibility of the alternative suggested, as contrasted to the impracticability of all other alternatives. These relationships have been systematically formulated and experimentally demonstrated by H. Peak and her colleagues.

The theory implicit in these persuasive techniques is that the individual will restructure his psychological field to maximise the attainment of his goals and values. Personality theorists have suggested that this theory is correct within limitations, namely, that when ego needs are threatened, the individual will respond to persuasive efforts by blocking and resistance. The resistance may result in more strongly reinforced attitudes than was true before the attempted persuasion. Again, the prediction is that the change attempts will have differential effects, depending upon the personal needs of the subject. Moreover, since the appeal to values operates more at the motivational than the perceptual level, the prediction is that for non-ego-defensive people there will be more change than in the first and second series of experiments. For the ego-defensive people there will be less change because of the phenomenon of resistance.

III. *Changing attitudes by attacking ego-defensive forces through the use of catharsis and direct interpretation*

Two techniques widely employed in psychotherapy which have possible application to group situations are permissive catharsis and direct interpretation. The basic objective is to help the individual attain insight and to restructure his attitudes accordingly.

1. *Permissive catharsis.* The free ventilation of thought and feeling in an accepting interpersonal atmosphere is an integral part of virtually all schools of psychotherapy. It is generally assumed that this type of expression helps the patient to change in the following ways:

(*a*) It offers him relief from the tension of burdensome affects such as guilt and hostility. The very act of expressing these feelings is supposed to drain off energy which would necessitate the formation of various somatic and psychic symptoms.

(*b*) It permits him to verbalise and hence objectify his inner conflicts. Such objectification leads to clearer self-perception. This clarification in turn makes it possible for him to utilise his intellect more effectively in choosing among alternative means of resolving his conflicts.

2. *Direct interpretation.* This technique uses a combination of logic and suggestion; patients are confronted with psychological explanations of their behaviour. These explanations are supposed to give them immediate insight into the internal factors (motivation) which determine their overt behaviour. The patient is then to bring the weight of his own common sense to bear upon the proffered insight and to change his behaviour in the light of this newly acquired knowledge. Thus, the lengthy process of uncovering is short-circuited and the patient is directly informed about the underlying motives of which he had previously been unaware. In employing this sort of approach, the therapist is generally obliged to rely upon the degree of authoritativeness with which the patient invests him and the sources (research findings, theory) on which he bases his interpretation.

It is granted, however, that these techniques will not be effective for all ego-defensive people. In some cases resistances may still be too strong to be affected, either by catharsis or by direct interpretation. The assumption is made, therefore, that people who already have enough insight to be concerned about their own conflicts and to be dissatisfied with some of their own behaviour

will respond to these techniques. Just as individual therapy can make little headway with patients who do not want to be helped, so group procedures will affect only those group members who are dissatisfied with their self-image, even though they do not understand the basis of their conflict.

In this type of experiment the direct interpretation can consist of psychological explanations of the dynamics of scapegoating and of displaced and repressed hostility. The objective will be to give self-insight. Individuals who are more insightful at the start will apply the interpretation to themselves more readily than individuals lacking insight. One advantage of the technique of direct interpretation is that it can be used in printed materials and is not confined to interpersonal communication.

SOME EXPERIMENTAL FINDINGS

(*a*) *Rational persuasion compared with authoritarian suggestion*

An experiment was devised by M. Wagman to test the proposition that different personality types would respond differently to various kinds of change procedure. It was postulated that non-authoritarian people would be more affected by rational persuasion than would authoritarian people, whereas the authoritarian personality would be more responsive to suggestion from power figures than would the non-authoritarian person. It was predicted that even when the influence exerted was in the direction of greater tolerance for minority groups, the authoritarian type of individual would be moved more by suggestion coming from authoritative sources than by a reasoned presentation of facts and arguments.

The measure of authoritarianism was the F scale developed by Frenkel-Brunswik and her colleagues, an instrument designed to give degrees of authoritarian character structure. The authoritarian suggestion consisted of statements attributed to power figures, which emphasised institutional values of efficiency and orderliness for their own sake, rather than as means to ends. The rational persuasion consisted of materials presenting facts and reasons based upon scientific study and historical and anthropological considerations of racial differences.

The subjects, 250 University of Michigan undergraduates, were given pre-measures at the first session, the influence attempt at a second session and post-measures at a third session. In the first session they were given not only the attitude scales, as measures of the dependent variable, but also the F scale and the Michigan Completion Test as measures of personality characteristics. In the second session, different influence treatments were used. Approximately 70 subjects received the authoritarian suggestion which was designed to move them in a less prejudiced direction; 70 subjects were given the rational persuasion materials, again directed towards movement towards less prejudice; a third group of 70 subjects was given the authoritarian suggestion designed to move them in a more prejudiced direction. Finally, 40 subjects served as a control group. In a third session, post-measures were taken which repeated the original attitude scales. The influence attempt was separated from the pre- and post-measures by the use of different experimenters and by structuring the situation as part of a different investigation.

The results of this experiment can be summarised as follows:

1. The rational persuasion materials were more effective in producing more favourable attitudes towards Negroes for the non-authoritarian than for the authoritarian subjects.
2. A number of the authoritarian subjects showed greater prejudice after the persuasion attempt than they had originally shown.
3. The authoritarian suggestion, when used to produce more favourable attitudes towards Negroes, had greater success for authoritarian subjects than did the rational persuasion.
4. The authoritarian suggestion, when used to produce less favourable attitudes, was more successful with the authoritarian than the non-authoritarian subjects.
5. A number of non-authoritarian subjects reacted negatively to the authoritarian suggestion designed to move them away from a liberal position, i.e. they became even more liberal.

(*b*) *Rational persuasion compared with self-insight*

The limitations of both prestige suggestion and rational persuasion are implied in the above experiment. These attempts are differentially effective and if used with certain types of people can actually

have negative effects. Moreover, authoritarian suggestion in one direction may produce change only as long as counter-suggestions do not occur. In a second experiment, rational persuasion was compared with the technique of direct interpretation.

The subjects were 300 women students at the Michigan State Normal College at Ypsilanti, divided into their own sorority and dormitory groupings. The experiment was in three stages. In stage one, all subjects took a battery of personality tests and filled out a number of measures designed to test their beliefs, feelings and attitudes towards relations with minority groups. The personality measures consisted of: (1) the Bender Gestalt test, (2) the F scale of Authoritarian Character Structure, (3) the Michigan Completion Test, scored for patterns of ego defensiveness, (4) a specially constructed thematic apperception test designed to get at hostility towards Negroes, and (5) fairly direct questions about the nature and kind of felt internal conflict. The attitude measures consisted of stereotyped statements about Negro characteristics, a revised Bogardus Social Distance scale, and twelve specific action issues. In addition, the subjects indicated their perception of the group position on two of the issues. In the second session, influence attempts to change attitudes were introduced and the attitude scales were repeated at the end of the session. In the third session some six weeks later, the attitude measures were administered at the start of the session, then different feedback procedures were employed, and finally some of the attitude scales were repeated before the groups were dismissed.

The primary change manipulation took place during the second session. Two different experimental procedures were employed. One consisted of the rational persuasion materials, very similar in nature to the experiment already described in this section. The purpose here was to give the individual an opportunity to restructure his beliefs in relation to factual and logical materials. The second influence procedure used on comparable groups of subjects employed the technique of direct interpretation. A case study was presented of a girl who had never acquired insight into her own hostilities and inadequacies and who developed in consequence certain social prejudices. The case was presented sympathetically and with sufficient similarity of background to the subjects so that they would have little difficulty in identifying with her. They were asked specifically after reading the case story to indicate how relevant they thought this analysis was to the problem of prejudice in general. They were also asked whether they found themselves employing any of the psychological mechanisms which the girl in the story unconsciously utilised. The rationale for this procedure was that it might give some degree of insight to girls who already were not wholly satisfied with their own behaviour. Thus, if they could see how all of us use defence mechanisms to buttress our own egos at the expense of other people, they might restructure some of their beliefs which have a defensive function.

The major purpose of the third session (some six weeks later) was to obtain a measure over time of stability of changes which had been produced in the second session. Since the girls had to be assembled for the taking of this test, their coming together was also used for trying different feedback procedures to produce further change. The measures of the dependent variable (the attitude scales) were given at the very beginning of the meeting. Then some groups were merely given a reinforcement of the original procedure they had been subjected to. Other groups were influenced further through the use of three group variables: (1) presentation of group norms, (2) a discussion of the treatment of Negroes in relation to the value of national security, (3) a similar discussion carried to the point of commitment to specific propositions about the problem. In both the second and third sessions measures were taken of perception of group consensus.

Though the analysis of all the data in this experiment has not been completed, these findings can be reported:

1. The interpretation materials produced more liberal attitudes towards Negroes than did the rational persuasion attempt.

2. The changes produced by direct interpretation were more permanent than were the changes produced by rational persuasion. After the six-week period there was more of a tendency to move back towards original position for people subjected to rational persuasion than for people subjected to direct interpretation.

3. The subjects who originally saw themselves as very much like the group persisted in estimating the group as close to their own position during the whole change process. This confirms T.

Newcomb's concept of a fairly constant tolerance by the individual for his own non-conformity.[5] People with a narrow limit either conform to changing group attitudes or autistically distort their perception of where the group is. And the most significant finding here is that this behaviour during the attitude experiments was related to personality measures of conformity obtained from the Michigan Completion Test and the California F scale.[6]

Notes

1. S. E. Asch, 'The Doctrine of Suggestion, Prestige and Imitation in Social Psychology', *Psychol. Rev.*, vol. 55 (1948) 250–76.

2. T. W. Adorno, E. Frenkel-Brunswik, E. J. Levinson and R. N. Sanford, *The Authoritarian Personality* (New York, 1950).

3. Attitudes towards Negroes were the dependent variable in our experimental programme.

4. E. R. Carlson, 'A Study of Attitude Change and Attitude Structure', Ph.D. dissertation, University of Michigan, 1953; M. J. Rosenberg, 'The Experimental Investigation of a Value Theory of Attitude Structure', Ph.D. dissertation, University of Michigan, 1953.

5. M. S. White, 'Attitude Change as Related to Perceived Group Consensus', Ph.D. dissertation, University of Michigan, 1953.

6. M. F. Barlow, 'Security and Group Approval as Value Systems Related to Attitude Change', Ph.D. dissertation, University of Michigan, 1953.

The forms of the family vary greatly between different societies. In the following extracts, three different variations are described. First the traditional family system of Japan, one which has undergone changes but still remains an important element of modern Japan; second the family among the Wodaabe Fulani, a cattle-herding people in northern Nigeria; and finally the contemporary British family.

The extracts illustrate a number of these differences. The typically small nuclear family of the British contrasts with the wider family groups of the Japanese and the Fulani, and the patterns of marriage which set up the family also differ in all three societies. There are differences too in the way the family functions in the context of the wider society. In the economic sphere, the Fulani family is closely tied up with their pastoral way of life. The Japanese family, on the other hand, has connections with the rural agricultural patterns, while the British family usually takes little direct part in production but has economic relevance 'as a unit of consumption, as a provider and improver of homes, as a possessor and maintainer of property'. In the political aspect of the family too there are differences. For the pastoral Fulani, relationships between families, often cemented by marriage and then the formation of the new family, have a political tinge. In the Japanese state, by contrast, the family has been 'the basic unit in both public and private local community organisation'. In Britain the family does not typically perform a direct political function, but it does play a part in the general functioning of government since it is the primary unit for much of the administration of the modern welfare state.

But besides these obvious differences, there are also underlying similarities between the three family systems. First, they share the same basic functions: socially sanctioned reproduction through the recognised formation of the family by marriage; the maintenance and socialisation of the children; and provision for the social welfare and control of the family members. Second, all three family systems must be seen in the context of the wider society. It is true that their detailed political or economic or religious functions vary, but the family always has some part to play in these spheres. It becomes clear that it would be impossible to understand the functions of the family in any of the examples discussed here if we looked only at the family itself and paid no attention to the structure of the society at large.

THE JAPANESE FAMILY

Ezra F. Vogel

THE IE,[1] PAST AND PRESENT

At the heart of Japanese thinking about the family is the concept of *ie* (usually translated simply as 'family'), a traditional concept which remains very powerful, especially in rural areas. According to the concept of *ie*, the present living family is what Zimmerman calls a 'trustee family' because it is only the guardian of the past and future generations which also belong to the *ie*.

Traditionally, one son, his wife and children were responsible for looking after family property and all other children left the *ie* to make their way elsewhere. The greatest part of the family property, the family treasures and graveyard were passed on, undivided, to a single son and his immediate family. The present head of the family was responsible for properly venerating the ancestors, for preserving the family wealth and standing and for controlling the behaviour of family members. Some higher-class families even had a code of law, a written code of customs to be followed by the family in all succeeding generations. The *ie* was also associated with various religious rites, and had a special alcove inside the house at which prayers were offered for the family dead and a family graveyard where the bones of ancestors were entombed and annual ceremonies conducted for the departed.

The living members of the *ie* felt a responsibility to their ancestors, and the most desirable and commonest way to fulfil this responsibility was to provide for the *ie* in the coming generation by producing a son or adopting a suitable heir and properly managing *ie* affairs so that its future prosperity was assured. To the extent that Japanese have believed in immortality, it had been an afterlife associated with the *ie* because it was thought that one existed after death only if one's spirits were cared for by the living members of the *ie*.

The form of the *ie* was well adapted to rural land conditions. For many centuries a heavy population has farmed the small

amount of arable land in rural Japan, and if the family plots had been subdivided, they would long ago have ceased to support a family. The *ie* system of passing the family plot down intact provided for family continuity without impoverishing the *ie* by constant subdivisions. Many who did not own their own land had such stable tenant relationships with their landlords or their lords that they were able to pass down the rights of use to their children and their children's children and so preserve their *ie*. In some cases the amount of land owned or worked by an *ie* changed, but this usually meant a modification of an *ie*'s prosperity rather than the beginning or end of an *ie*. In comparison to most countries, there has been a striking continuity in ownership and use of land by a given family. In the case of the small business family, this practice of passing things down to a single heir meant that accumulated capital was not easily dissipated by division among heirs. Many small businesses, particularly in certain special crafts, have had a continuity within a given family comparable to the continuity in landownership.

The head of the *ie* had great power over the behaviour of its members, although in certain powerful families there was a family council which shared in important decisions. Because the household head controlled the inheritance, he decided if the first son deserved to become the heir or whether the headship should pass to a different son. Ordinarily this decision was made while the children were still fairly young, and the heir (usually the first son) was given preferential treatment even before he actually became the head of the family.

Because family continuity was of paramount importance, a family which had no children adopted a child. If a family had a daughter but no sons, the man (probably a younger son) who married the daughter ordinarily would take over the family name and continue the *ie*. If the family had no children, a son would be adopted, and because the *ie* was closely tied with the family enterprise the young man was usually selected largely on the basis of his ability to continue the family enterprise.

The *ie* has long been the basic unit in both public and private local community organisation. Ordinarily each *ie* (household) had one representative in the village councils, and even private organisations like the local women's club had only one representative from a household. In community matters it was not the individual who was the unit, but the household, the *ie*. This provided continuity to community organisations and made planning and control of community activities possible. Each household head was responsible for seeing that members of his *ie* abided by the decisions of the community councils and government agencies. Although the power of various *ie* has risen and fallen, there has been considerable stability not only in the number of *ie* in any given rural community but in the relative status of each within the community. Rural land reform after World War II did change the traditional power structure of the villages, but the number of *ie* in a particular village has remained relatively constant.

Ordinarily the first son inherited the family property, but if the household head felt he were unsuited for the job as irresponsible, another son or an adopted son would be assigned the role. Generally, children were trained in their youth for their respective roles of inheriting or leaving. Grandparents even distinguished between *uchi mago* (children born to their heir) and *soto mago* (other grandchildren). Kinship terminology centres around the *ie*. A person was ordinarily called by his position in the *ie*: *shujin* ('household head'), *okusan* ('wife of household head'), *choonan* ('first son'), *choojoo* ('first daughter') and so on. Although the firsts on was almost invariably called by his rank (even as a baby he was commonly called *anchan* ['little first son']), children who were later to leave the *ie* were often called by their personal names or by nicknames.

The young wife of the first son was commonly referred to as *uchi no yome* ('our *ie*'s bride'), for she was not only the bride of her husband, but the bride of the entire family. She was treated in many ways as a daughter of the household. Divorce consisted essentially of her being sent back to her original *ie*. Just as the *ie* had the major role in deciding which bride would be selected for the family, so they decided whether or not the young bride should be returned to her family because of negligence, poor attitude or failure to produce a son.

Traditionally, the *ie* was responsible for the welfare of all its members. The heir and his wife cared for the elderly as a matter of course, but their duties extended beyond that to any member who became sick or lost his job or whose marriage failed. Although the *ie* might not be pleased to welcome them back since facilities were

limited, they nevertheless felt responsible for the welfare of former members.

The concept of *ie* is still strong in rural areas and in certain special crafts where skills are passed down from generation to generation, as the kabuki theatre, or in independent professions like medicine. In such cases, the concept of *ie* is still very closely connected with the method of economic livelihood. The children inherit from the parents the property or the skills which provide a substantial help in the family's livelihood. However, in large cities where men work for large factories, business firms or government offices, bureaucratic positions are out of the control of the *ie*, and inheritance is much less important than obtaining a good job in a respectable organisation.

After World War II, as part of the attempt to democratise family relations, the revised family code provided that there should be equal inheritance among the children. Although this law has usually not created such serious difficulties in the city where family property is in more liquid form and inheritance is not critical for earning a livelihood, in rural areas it creates the serious problem of uneconomic land division. To avoid this danger, farm families have resorted to informal techniques, such as requiring other children to give land to the heir as gifts, in order to preserve the family land. The pattern of single inheritance thus remains strong in rural areas despite the provisions in the new Civil Code. In the urban areas inheritance is now divided, but there is not yet a broad base of welfare services supplied by the government. Since the *ie* has declined and no longer provides for family members, there is a serious problem.

In the past hundred years, the Confucian ideology was strongly supported and widely publicised by the national government, which tied it in with loyalty to the *ie* and the government. Although no longer the official ideology since the war, Confucian ideals are remembered by all adults, many of whom committed them to memory when in school. The official Confucian ideology required that the child be totally obedient to his parents, the wife be totally obedient to her husband, and when he dies, to her son, and that the younger brother be totally obedient to his elder brother. Although this philosophy imposes a strictly ordered and disciplined society, in practice there was a considerable freedom. For example, a young bride was supposedly a subservient creature, completely obedient to her new family, controlled in her behaviour and quick to offer help and assistance. However, once she had produced a male child or once her mother-in-law died and she became the lady of the house, she had a great deal of power in the home. There was a sharp division of labour between men and women, and since the men did not participate at all in household work, women had considerable independence in managing their affairs. In addition, women generally managed the household finances, a practice which gave them far more authority than the official Confucian ideology of obedience implied.

THE RELATIONS BETWEEN STEM AND BRANCH FAMILY

The *ie* took considerable care in planning the futures of their excess sons and the marriages of their daughters. All prospects were studied thoroughly to determine the best placement the family could command. In some cases, there were opportunities for the excess sons to marry into families which had no heirs. In other cases, small businesses which had grown up in the countryside absorbed many local excess sons. Sometimes they took on a position of *kobun* (a kind of ritual sonhood) in relation to the *oyabun* (the boss who was a kind of ritual father). In certain parts of Japan new land was available or a few plots were large enough to permit division among the sons. During the era of Japan's overseas empire, excess sons often were sent to colonial areas. If finances allowed, the second son might be given money by the *ie* to start a business of his own. In recent times, the most common pattern has been to send the second son to a town or city to earn his own livelihood. The *ie* would generally make all work and living arrangements beforehand through acquaintances or acquaintances of acquaintances.

If the migrant did not join another *ie*, he might remain an official member of his original *ie*, and his family would officially be considered a *bunke* ('branch family') of the *honke* ('main family'). In the rural areas and in certain business areas, the relationship between main and branch family was bound in common economic enterprise. In such cases, the branch family was typically firmly under the control of the main family; many times the branch family was established with capital lent by the main family and remained in business thanks to its assistance. Even after two or

three generations, the descendants of the branch family were expected to be loyal and obedient to the descendants of the main family. The relationships between the stem and branch families were expected to be much the same as they had been in previous generations.

The extended family group which unites stem and branch families is called the *doozoku*. Many interesting cases of *doozoku* relationships have been discovered by Japanese anthropologists and sociologists. In some cases, the *doozoku* became a fairly large organisation with a great deal of power. In many new business organisations, the relationship was developed along the lines of the stem and the branch. For example, if the owners of a successful drugstore wanted to start a new branch, and if they had a competent second son, they would give him the management of the new store. This was the pattern of growth for many small shops in the past few hundred years. When the family did not have a second son, it could adopt a very trusted assistant and establish him independently as, in effect, a branch family. He was permitted some independence in running his business, although the major decisions would be made by the main family. Often the branch family sent its children to the stem family for a period of apprenticeship, to keep close liaison between the families. In some large businesses and in certain large successful farm families, the *doozoku* was the core of the organisation. The relationships among the various people in the business were determined by their family position: whether they were in the main family or the branch family. Thus family position had a great deal to do with determining the precise position and power of each person in the organisation. At the same time, however, decisions on the formation of branch families were influenced by the competence of the people involved and the business opportunity. Questions of inheritance also were governed in many important ways by questions of ability. If an only son in the stem family were considered incompetent, a more capable son from a branch family would be made the heir of the stem family.

There is disagreement, even among Japanese scholars, as to how widespread and how significant the *doozoku* pattern was. In many cases it was obviously important, but no quantitative data are available from previous times. All agree, however, that in contemporary Japan, with few notable exceptions, the *doozoku* is merely a vestigial form with little significance. Even in rural areas where people are aware of their *doozoku* relationships, the *doozoku* has little power over its members. A major cause of its decline has been the process of industrialisation and urbanisation which has shifted power from the stem to the branch families. The main family remained on the farm or in the traditional occupations, while the excess sons went into the new industries which have profited most from modern economic developments. Today urban families are more prosperous and more powerful than their rural relatives. Hence, the branch family which is theoretically subordinate has typically grown in relative power to the main family, and this shift has undoubtedly hastened the decline of the *doozoku*.

Examples of weakened *doozoku* ties are numerous. It was impossible to demand that a second son in the city, earning more than the farm family, continue to subordinate his wishes to those of the main family. In many of the richer rural families, the contrast was even more striking. After World War II when many large landholdings were broken up, elder sons even among well-to-do élites found themselves impoverished compared to younger brothers who had established successful businesses.

Although the younger son who moves to a city may go through the form of establishing a branch family, the bonds between his branch family and the main family are either weak or non-existent. When he first migrates he may ask his family's permission to get married and he may follow their suggestions for a bride, but once he and his wife are established in the city, they are virtually independent and feel little obligation to the main family. With the rise of urban prosperity in relation to the rural areas many stem families now seek the help of their branch families in placing their children in schools and jobs. During the war-time air raids and food shortages, city residents turned to their rural stem families for help, but today requests for favours and assistance generally flow from farm to city.

MARRIAGE

Marriages in Japan have traditionally been arranged by families with the aid of go-betweens. Where the children married family or childhood acquaintances, formal meetings were unnecessary, but the usual pattern was to hold a *miai* (a meeting of the young

man and the eligible young lady and other representatives of their families). Before the *miai*, both families discussed the matter thoroughly and quietly investigated each other through friends and, in some cases, private detectives. When the investigations and discussions were favourably concluded, the formal *miai* was held. If the ceremony went well, an engagement and marriage would follow within a few months. If not, another *miai* with another family would take place. Ordinarily, the number of *miai* that any person would go through was very limited, rarely more than three or four. If a girl went through this many *miai*, she and her family would acquire a reputation of being too particular and would find it difficult to make new arrangements. Because there was not a diffuse close relationship between the couple, their families relied on objective criteria such as the family's status and wealth, the health, life expectancy and strength of the family line, the appearance of the girl and the economic prospects of the young man. Recently, they rely increasingly on level of education and the quality of the school as a basis for judgement. In rural areas, there is still a brief courtship period at most, but a series of carefully regulated meetings is often the only kind of premarital contact. Marriage arrangements are commonly negotiated through go-betweens who confer with both families in order to satisfy their respective demands.

At the time of engagement, the 'receiving family' (ordinarily the young man's family, but in the case of the adopted son-in-law the wife's family) gives a present of money to the 'giving family', and this amount (usually determined by the go-betweens) is used to pay for part of the dowry which the new member of the family brings to her new family. The wife who marries into her husband's family ordinarily brings her own clothing and enough furniture to last her and her husband for many years. In some rural areas it was the custom for the bride to display the things she was bringing to the marriage to all the neighbours so they could see the quality and size of her dowry. The wedding ceremonies were paid for partly by the bride and partly by the groom, and in most cases required a very large expenditure. Since World War II there have been various movements to simplify wedding ceremonies in order to cut down the expense and the lavish display of conspicuous consumption, but on the whole the movement has not been very successful and wedding ceremonies are still expensive, especially for the girl's family. Many girls work from the time they leave school until their marriages to earn enough money for their dowries.

Although most married couples were married by *miai*, it is increasingly common in urban areas for the couple to meet and form their own relationship on the basis of love (*renai*). It is still usual, however, for parents and child to make the decision about marriage jointly. Even if parents take the initiative in arranging the marriage, they must obtain the consent and approval of their child before a final decision. Most families prefer their children to meet through proper channels and conduct proper investigations before the children are allowed to develop relationships of their own. Parents are not ideologically committed to the pattern of the *miai*, but they feel that proper arrangements give a marriage a better chance of success and help prevent heartbreak on the part of young people who lack the experience and emotional control to make wise choices. Although most children now prefer to have *renai* without *miai*, it is common even in city areas for a family to investigate a prospective spouse and his family, to have the *miai* and then allow the couple to meet alone a number of times to decide whether they wish to marry or not. Although young people are permitted a period of *renai* after a *miai*, it is still preferred that a decision be made after a few meetings and that engagement and marriage ceremonies be held quickly.

Although young people occasionally meet on their own, it is still considered somewhat reckless to form a friendship without some kind of an introduction. Increasingly, however, siblings and classmates are replacing parents as sources of introduction, and in some cases, especially at co-educational schools and places of work, young men and women develop friendships without any special introductions.

Even in *renai* marriages, the courtship period is so brief that a couple's affectional ties commonly develop mostly after marriage. In rural areas young people are not expected to have a diffuse romantic relationship before marriage; it is assumed that after marriage the couple will gradually get to like each other and feel more comfortable with each other. The greatest hindrance to marital adjustment is the lack of privacy. Many young rural couples have little time to be alone, especially if they live with grandparents. Some do not ordinarily have private sleeping

quarters. Even in urban areas, privacy is limited, for the youngest child often sleeps on a mattress between his parents until he is of school age. Nevertheless, it is expected that over a period of years the couple will become more affectionate.

Until prostitution was declared illegal shortly after World War II, it was fairly common for young men to have sexual experiences before marriage, and in fact it was considered desirable for the husband to have sexual knowledge before marriage. The woman, however, was a virgin until marriage and in some cases she had no sexual information until her wedding day when she was given a lecture and a brief explanation by her mother, the go-between or a close female relative. Now sexual education is more widely available in the schools and in regular fairly widely read publications. The studies of Japanese sexual life suggest that there is little satisfaction, particularly on the part of women, from sexual relationships early in marriage, but that after several years the level of satisfaction increases greatly. These studies also suggest that there is relatively little foreplay or afterplay in comparison to sexual relations in this country, as reported by the Kinsey studies. Although some men have extra-marital relationships with single girls they meet at places of work or entertainment, it is virtually unknown at most class levels for a married woman to have any kind of friendship with a man other than her husband.

Close affective relationships with parents continue much longer than in the United States so that it is not uncommon for the young man to have closer affectional ties to his mother than to his wife, particularly early in marriage, and for a young bride to feel more love for her parents than for her husband in the early years of marriage. When they live in the same household, a bride and her mother-in-law often have serious disputes. Families make great efforts to avoid this situation for it creates one of the most crucial conflicts in family relationships. Yet if the mother is widowed there may be no other alternative that is compatible with feelings of kindness towards the elders. Traditionally, the husband sided with his mother, and the wife acquiesced, but as the wife's power has increased, the relationships have become more equal and the rivalry more open and severe. Even though the husband generally tries to stay outside this conflict, it still commonly imposes a severe strain on the marriage. Because the affective relationship with one's parents is not neutralised at such an early age or as completely as in America, it takes much longer to form close affectional ties between husband and wife, and the presence of a parent in the household can impede the progress of marital affection. With the period of courtship contact increasing in intensity, the affectional relationship before marriage is stronger, but the new ideal of love before marriage is not always realised.

INTERNAL FAMILY FUNCTIONS

Adaptive

Traditionally, the major economic functions were performed within the *ie*. In the farm family, and in small merchant families also, wives and children often worked with the father in carrying on the family enterprise. To some extent, such families tend to remain an economic unit, and much of the hard work is performed by the females as well as by the males. If the father should become sick or die at an early age, his wife or the children, if old enough, would commonly take over the family enterprise.

In the earlier stages of industrialisation, it was common for women to take on work not associated with the family enterprise in two ways. One way was going out to work in a factory. If the factory was not close by, then the women could live in a dormitory. This pattern, however, did not ordinarily apply to married women. Ordinarily young girls would go to live in the dormitory under fairly difficult economic circumstances, and when they had accumulated enough money for a dowry, they returned home to be married. The second pattern was for the woman to take work into the home. Small work like sewing, weaving and some kinds of small craftwork could be let out to women. The latter pattern applied to married women. Thus it was very rare for any married woman to perform work outside the family. Once married, the wife's primary assignment was in the home, and she did not have to separate her home from her work. Only if she were widowed at an early age would a woman enter the labour force, and even then the preference has been for her to set up a small shop or offer some kind of service like hairdressing or teaching young girls how to prepare to become housewives, whereby she could do the work in her own home.

Just as the wife ordinarily did no work outside the home, so the husband ordinarily did no housework. Even today husbands

rarely participate in housework, even if the family enterprise is connected with the home and the wife works alongside her husband in the family enterprise. The division of labour between husband and wife typically remains very sharp except for that common sphere in the case of a family enterprise.

Women working outside the home remain a very small part of the labour force, and though they have taken on many different kinds of jobs, the common pattern remains for girls to work in various places when they are young, accumulate some savings for their marriage and then quit work upon marriage. Except for volunteer philanthropic work among upper classes or employment of wives in economically deprived families or of widows and divorcees, it is still very unusual for women to work outside the home. Hence, it is normal only in those families with a family enterprise in which the wife performs major economic activities.

However, in many families with small enterprises, especially independent farms, it has become increasingly common for the husband to take on part-time outside employment. More than half the farmers of Japan now have some such side work in addition to farming. This means that much of the work and management of these small independent economic units is passed to the women and children.

Where these independent economic units are of significant size, and inheritance is adequate to provide economic activity for a child, the family enterprise is generally passed down to one child, maintaining the traditional *ie* pattern. In the case of small shopkeepers where the capital involved is small, there is generally no point to passing down the family enterprise, and the economic unit is the nuclear family or household rather than the traditional *ie*. With the growth of large-scale industry, however, it has been increasingly common for the man to work outside the home and for all major economic activities to be performed outside the home.

Until the end of World War II, the standard of living was generally so low that relatively little money was available for consumption, and the opportunities for purchasing were often circumscribed by regular store–customer relations. Since the war, however, with the rise in the standard of living and the increase in quantity and variety of consumer goods, consumption has become a major economic activity of the family. However, because the standard of living is still limited in comparison with that in America, for example, great care is exerted in maintaining the supplies a family has at its disposal.

Co-ordinative

Most family activities are determined by custom and do not require active decision-making on the part of the husband and wife, except to reaffirm or rebel against customary practice. In case of conflict or disagreement between husband and wife, the husband has traditionally been able to have his say. According to traditional ideology the wife was supposed to obey without question. In actual practice, however, the typical husband has taken relatively little interest in household activities, and the wife, albeit quietly and subtly, moved in to fill the vacuum. Even though commonly deferent to her husband, the wife may be very effective in seeing that her wishes are carried out. Ordinarily, for example, expenses are handled by the wife, though husband and wife do consult on major issues.

Integrative

Ordinarily, there is a very close solidarity between family members, especially *vis-à-vis* the outside. Generally, the home is a place where everyone can relax, in sharp contrast to the greater formalities and politeness required in many parts of the outside world. However, a woman ordinarily has few independent means of acquiring more facilities for the home, for herself and her children except by making requests to her husband. However subtly she may make such requests, the husband may feel the pressure and find it difficult to relax much at home even though the wife and children may be eagerly trying to attend to his whims and pleasures. As a result, he may turn to the outside for recreation and tension release, while the wife and children, whose place is in the home, have their recreation (especially talk, radio, television) at home.

Rarely do husband and wife go out as a couple, and rarely (except in wealthier upper-class families) does an entire family travel together. The wife may go out with a group of ladies for an outing, and the children will have an opportunity to go on school outings and trips, but family 'togetherness' is generally limited to the home.

CHILD TRAINING AND SOCIALISATION

As in other countries, the family size has diminished rapidly with modernisation. With the rapid industrialisation, families have shown a desire to provide their children with adequate education and a variety of consumer goods so that they have been motivated to have only a small number of children. However, in places where the concept of *ie* has been strong, the family is highly motivated to produce an heir. In fact, the main purpose of marriage was to produce an heir for the *ie*. In a previous generation, with the high rate of mortality, it was common practice for a family to produce at least two sons, that one might take over the *ie* in case of death of the other. Even today, many families say they would like to have three children, two boys and one girl, in order to be sure that they have an heir. With improved health measures and increased longevity, however, there is no longer the urgency of producing excess children to ensure *ie* continuity. For branch families or modern families where the power of the *ie* is weakened, a family is often content with having one or two children, even if there is no boy. Once a certain stage of modernisation is reached, and the family is motivated to have a small family, it seems to make little difference whether the method of birth control is abortion or contraceptives. Abortion is legalised and still very widespread. About as many pregnancies end in abortion as in live births. But in those areas where special campaigns have been conducted to teach the use of contraceptives, there has been no particular objection to contraceptives, and they have often replaced abortion.

Although the contemporary family tends to be small, the wife ordinarily devotes herself completely to her children. She takes on no outside work, and her social activities are typically limited to the immediate neighbourhood and to meetings connected with the local school. As a result, her entire life centres around her children. The wife of a small shopkeeper or farmer who assists her husband and also is completely responsible for the care of her children carries a double burden which the urban new middle-class housewife does not envy. Unlike many of her American counterparts, she has no desire to work outside the home and is very pleased to be able to devote all her time to house and children. Generally, the care of children involves not only looking after their physical needs, providing basic socialisation and handling emotional problems, but also includes the responsibility of assisting them with homework and giving assistance in arranging their marriages and placing them in appropriate jobs. Parents are responsible for and direct their children's activities to a greater extent and for a longer time than in America.

The mother is much closer to the children than in the comparable American family. When a child is young he is almost never away from his mother; she will even carry him on her back while working or shopping. She generally nurses her child until he is a year old, and mother and child often bathe together until the child is of grade-school age. She generally sleeps on the same mat with the youngest child until he is at least three or four years old, and even when older the child may sleep with father or a grandparent, or at least have his own mat immediately next to the parents.

The emotional relationship between mother and children is correspondingly close. The child, even at high school age, spends nearly all his evenings at home and has much closer relationships with his mother than does the comparable American child. Fathers who work outside the home generally work very long hours and spend less time at home than American men so that family activities are left more completely to the mother. The father usually does little disciplining, and his occasional demands on the wife and children are generally moderate. (The exceptions to this description are farm or small shop families where the father works at home. In those cases the father supervises family activities in greater detail. Since the proportion of fathers in such occupations is now declining, this pattern is less widespread than previously. Even in rural areas, most husbands now have side jobs away from the farm, leaving farm work to wives and children, except during the busiest season.) In most middle-class families, however, the wife and children have considerable independence from the father, and the mother has considerable authority in handling daily activities, household management and family finance. As in most societies children learn to take on attitudes and behaviour expected of them by their parents simply because no alternatives are presented, and parental attitudes are sufficiently clear and unconflicting. Most children feel that their mothers are sacrificing for them and it is only fitting that they follow their mother's

wishes. To the extent that more serious sanctions are invoked they generally involve the threat of isolation and separation.

When the husband's mother lives with the family, there is often disagreement between mother-in-law and daughter-in-law about the rearing of the children. One method of easing the tension is for the grandmother to assume the major responsibility for one of the children and for the mother to have primary responsibility for the others. In fact, the child assigned to the grandmother is often literally called 'grandmother's child'. If, for example, a child is about three when a new baby is born, the three-year-old may be assigned to his grandmother, leaving the mother free to care for the infant. The grandmother's child will then sleep with the grandmother, take baths with her and be cared for by her. In turn he will run errands and do little favours for her. Ordinarily the grandmother shows marked preference for 'her child', and she will stand up for his rights *vis-à-vis* the mother.

Overt discipline such as spanking and hitting is rare among Japanese families, and because the mother–child relationship is so close, the mother can rely on indications of approval and disapproval to control behaviour. The techniques of child-rearing include abundant use of praise, warm support, constant attention to needs, creating fear of the outside and vague threats of withdrawal of love. All these techniques tend to create a very intimate mother–child relationship and make the child responsive to the mother's wishes. A Japanese mother can tolerate and in fact encourages much more dependence on the part of her children than an American mother does. She also indicates very clearly how the children should behave and they are quick to learn what is expected of them. Many mothers and children do not regard their training as discipline because it does not involve the invocation of principles, 'teaching children lessons' or physical punishment. But children are trained just as effectively, though more subtly handled, than children raised by the rod.

Deviance

On the whole, the Japanese family has retained its power over its members, and the amount of deviance is extremely small. Japan is one of the few countries where the divorce rate has actually gone down with urbanisation. Today, about one in ten marriages end in divorce, which is about one-third the rate of 1885. The previous high divorce rate was related to a pattern of the *ie* sending back the bride if she was not satisfactory, and her original *ie* then felt the responsibility of looking after her. With the weakened *ie*, there may now be no such unit to look after the divorced woman. The low divorce rate of today seems to be intimately related to the general low mobility in other sectors of Japanese society; men rarely change jobs, and families rarely change residences. The alternative opportunities for women for employment and remarriage are so limited that women usually make every effort to adjust to their marriage.

Various acts of delinquency and crime have increased somewhat over the last generation or so, but the increase has not been of major proportions. Until very recently, there have been virtually no adolescent gang activities, because children have had little opportunity for activities outside the family. Virtually all their free time was spent in the home. Recently, with increased freedom, there has been an emergence of some teenage gang activity, but it is still much less widespread than that in the United States.

By and large, the Japanese family has been and remains a stable unit, with adequate control over its members, even though the household unit is increasingly limited to the nuclear family members.

Note

1. The word *ie* can also be used to mean house or household, but in his chapter it will be used in its meaning as a 'lineage group'.

THE WOƊAAƁE[1] TODAY: FOUNDING A FAMILY AND A HERD

D. J. Stenning

We have given an outline of certain features of the social organisation of the Woɖaaɓe of West Bornu as we understand it to have been in the past. We have throughout taken the viewpoint that Woɖaaɓe, like all Pastoral Fulani, have been affected by historical events which were not initiated by them, but in which they played a part, or to which they had to accommodate themselves. The cumulative effect of these changes has been to draw them

more closely into the administrative structure of the States of the Western Sudan, in this case the eastern Fulani Emirates and the State of Bornu. This has been accompanied by their greater participation in the cultural world of Islam in the Western Sudan; in its rituals, its ceremonies and its vocabulary.

In this process the traditional institutions of the Woɖaaɓe have been modified. We have noted the effects of their participation in the later stages of the *Jihad*[2] in north-eastern Nigeria, in which they acquired slaves. We have seen that conditions in the eastern Fulani States prompted them to seek a new habitat, and that the dual allegiance which this involved caused an enlargement of their ideas of leadership. We have suggested that Islamic influences have been at work in the ritual sphere and, to some extent, in their notions of social control. We have noted the disruption caused by invasion, natural calamities and the early administrative efforts of the British. We have seen that certain of their chiefs were given a new role, the tribute was transformed into tax, and that a form of settlement was instituted. But the Woɖaaɓe recoiled from the idea of any wider political unity, and remain today a minority in an alien State. They pursue their pastoral interests and express the values associated with them not only by exploiting the natural habitat, but by adjusting themselves to the demands of a wider social environment.

In this part, we look more closely at Woɖaaɓe social organisation to see to what extent the institutions they believe most necessary for their pastoral life are also adapted to the exigencies of their participation in the life of the Islamic Western Sudan and to the demands of the State under British tutelage.

INTERDEPENDENCE OF FAMILY AND HERD

Our first concern must be with the family – the social group hitherto described as the basic economic unit of Woɖaaɓe social organisation.

The seasonal movements of the Woɖaaɓe, which will be described more fully later, take full advantage of fluctuations of pasture and water, and allow for emergencies caused by disease and injury to stock. A flexible organisation of the groups undertaking these movements is therefore demanded. Thus a high degree of independence in the choice and exploitation of grass and water supplies is accorded to small groupings. It is the household based on the simple or compound family which enjoys the right of seasonal deployment of the herd of cattle it owns. The analysis of the household, and its formation within the homestead it inhabits with its herd, demonstrates the relationship of these human and bovine units.

A herd we define as a distinct group of cattle consisting of at least one serving bull and a complement of cows, heifers, calves and pack-oxen. It is distinct in that it normally inbreeds, spends the daytime period of grazing apart from other similar herds, is watered separately and spends the night in a corral reserved or constructed for it and marked by its special smudge fire. The corral is part of the homestead of the male herd-owner, who is responsible for its day-to-day management.

In the homestead of the male herd-owner live his household; normally a simple family composed of himself, one wife and her children, or a compound family composed of himself, his several wives and their children. In circumstances in which cattle are the sole basis of subsistence for such a household, meat does not form a regular or staple diet and animals are not often sold, since this represents a withdrawal from the capital stock of the herd. Killings are confined to male beasts or, where possible, sick or maimed animals; such killings occur only upon ceremonial occasions. Sales take place only when there is an overriding need for cash, for example, to buy corn in a bad dry season or to raise tax. In these circumstances the family lives on milk, for, unlike some East African pastoralists, Fulani do not drink cattle blood or make blood foods. Either milk or milk products must be drunk or eaten *ad nauseam* or milk must be sold or exchanged for other foods.

Thus the limiting factor in the dependence of a given family group upon its herd lies in the milk output. If the milk output falls temporarily below a certain minimum, cattle will have to be sold to pay for food, thus potentially reducing still further the fecundity of the herd. Humans share milk with calves; if they make too great a demand upon the milk of the herd, the well-being or even the lives of the calves, and therefore the future fecundity of the herd, is prejudiced.

Good husbandry for a Fulani herd-owner thus involves maintaining a milk yield sufficient to support his dependants at all seasons

Lactation must not only be adequate, it must be continuous. Since lactation is dependent upon the birth of calves, the main interest of the herd-owner is in a steady yearly increase in his herd.

A given milk yield might be maintained by controlled breeding programmes, the plantation of certain types of pasture, controlled feeding and watering – in fact, the common apparatus of dairy science as it is practised in more advanced societies. Like all Pastoral Fulani, the Boɗaaɗo[3] cannot adopt such a course. He finds himself in a situation of seasonal shortages of water and pasture and severe endemic diseases, in harsh tropical conditions in which the maintenance of the herd in the state required merely for continual subsistence may be achieved only by constant seasonal movement and periodic migration. Cattle cannot here be thought of as milk producers only. They have to be capable of long miles of trek to, from and between pastures, and able to undergo seasonal irregularities in both the amount and type of fodder and water. The Boɗaaɗo herd-owner therefore maintains his milk yield, not by a dairy technique directed to realising the milking potentialities of individual animals, but by one in which, for the individual animal, milk yield is not the only desirable property.

We have here been considering the herd as supporting the human family dependent upon it; we have taken the family as 'given' and examined the general properties of the herd associated with it. But there is a sense in which the human family supports the herd. Fulani cattle are not natural groups of wild animals followed and exploited intermittently by humans. Fulani herds are domesticated in a particular way, and this domestication entails a degree of special organisation in the families dependent upon them. Desirable pastures have to be sought and cattle led to them. Water supplies have to be arranged and cattle watered regularly. Diseases and accidents of many kinds have to be avoided, or their results treated. Cattle have to be protected from the attacks of wild beasts, particularly hyena, and thefts must be prevented. Birth of calves has to be assisted, for Pastoral Fulani believe that a calf born in the bush brings ill luck on the herd, and should be sold on weaning. The supply of milk available to calves and humans has to be controlled by careful milking. Among the Woɗaaɓe the division of these tasks between the sexes is clear-cut. Men have to do with cattle, their seasonal movements, daily pasturing and watering and veterinary care. Women have to do with milk and its marketing, in addition to their domestic tasks of food preparation and the care of the homestead both at rest and on the move. Adult men are herd-owners and managers, male children and adolescents are herdsmen. Adult women are dairywomen and purveyors of milk, female children and adolescents are dairymaids. The Woɗaaɓe family is a herd-owning and milk-selling enterprise.

Given this strict division of labour, and a herd of a certain size, a herd-owner's family must attain a size commensurate with its responsibilities towards its herd. It must also maintain a balance of the sexes, so that these responsibilities may be efficiently carried out by appropriate members of the family. A Boɗaaɗo herd-owner must have at least one wife. Man and wife must reproduce at a certain minimum rate and in a certain proportion of potential milkmaids and herdsmen. When the size and increase of the herd and of the family are adequate for the control of the former and the subsistence of the latter, then family and herd may be said to be in equilibrium. In the life history of a family, expressed as the lifetime of the male household head, such a state of equilibrium is never permanent; a whole range of temporary, partial or potential disequilibria arises which requires resolution. One party to the disequilibrium – the cattle – can do nothing to adjust the state of affairs, though Woɗaaɓe quite reasonably believe they may approve or disapprove of certain acts of their owners. The other party – the humans – can and do take action both on the cattle and themselves to readjust the relationship of family and herd.

The herd and the family are brought into existence by first marriage by betrothal, the procedure and symbolism of which demonstrate the interdependence of family and herd. The maintenance and increase of herds is effected largely by seasonal movement and by co-operation between herd-owners. The adjustment of family size and composition is carried out by subsequent monogamous marriages, polygynous marriages and the divorces which these may entail.

THE HOMESTEAD AND THE FAMILY

The orientation and arrangement of the Woɗaaɓe homestead symbolise the crucial relationships within the family and also that of the family to its herd.

Compared with the clusters of circular thatched mud huts which in some form or another constitute the hamlets and villages of most of the sedentary populations of the Western Sudan, the homesteads and camps of the Woɖaaɓe seem haphazard and rudimentary. In the wet season even a large camp blends with the bush, and in the dry season it is possible to pass within a few yards of a homestead without realising it is there. Household equipment is limited to the amount which may be carried on the head or on pack-oxen, and shelters must be made of whatever tree foliage the district has to offer.

Woɖaaɓe are described by their sedentary neighbours as 'the wind which blows through the bush; when they have gone, all that remains is a few stripped branches and a patch of bare earth'. The pastoral Boɖaaɗo has no sense of attachment to a specific plot of land he calls his home. Rather, he is at home in any of the favourite pastures where he or his father have tended their herds, in any of the markets where his womenfolk sell milk, and on the cattle tracks (*burti*, sing. *burtol*) and footpaths (*laabi*, sing. *laawol*) which join them.

Nevertheless, the homestead, though impermanent and simple in the extreme, exhibits important elements of formality and uniformity.

Woɖaaɓe homesteads and camps always face west (see Fig. 9.1). In the context of the homestead, west is synonymous with 'front', and east with 'back'. In Fulfulɗe,[4] north and south are used for the left and right sides of the homestead respectively, as seen by an observer facing the homestead and looking east. In all but overnight camps a curved back-fence of branches cut from nearby trees (*hoggo*) is put up to ensure a measure of privacy, keep out hyena and deflect the course of stampeding cattle frightened by their nightly visits in the wet season.

Immediately in front of the back fence are set the beds of the household (*leese*, sing. *leeso*). Four stout forked stakes (*daggaaje*, sing. *daggaaral*) are driven into the ground to form the corners of a rectangle 3 ft by 8 ft. Poles (*sakkaaje*, sing. *sakkal*) of about 4 in. in diameter, which are part of the domestic equipment, are placed on the stakes to form the two long sides of the rectangle. Thirty of forty special sticks (*sammbaaje*, sing. *sammbal*), an inch or so in diameter, are laid crosswise upon them; these may be adjusted to fit the curvature of the body in repose. One or two basketry mats (*daage*, sing. *daago*) serve as mattress, bark-fibre mats (*ndampi*, sing. *ndampa*) as blankets, and bark-fibre chafing pads from the pack-bull's harness are used as pillows. The shelter for this bed (*garba*) is a structure similar to it but on longer uprights (*sompooje*), forming a platform on which are placed the household utensils. In the wet season the small cross-poles of this structure are extended as a framework for additional basketry and bark-fibre mats which provide some shelter from the rain.

Alongside or in front of these bed-shelters there may be other ancillary shelters put up for specific purposes. These are low, beehive huts of springy boughs covered with bark-fibre matting (*bukkaaru*, pl. *bukkaaji*) or tall shelters of heavier, leafy boughs (*suura*). These shelters do not always contain a bed made with poles and stakes, but usually a rough couch made of bundled grass and bark-fibre mats. In front of the bed-shelters there are domestic hearths (*hiite suudu*), which consist simply of three stones or parts of white-ant nests, forming a support for a cooking pot, supplemented perhaps by an iron tripod bought in the market. The bed-shelter, with its hearth, ancillary huts and domestic area, is known as *suudu* (pl. *suudi*).

Near these shelters and in front of them is kept the household's stock other than its herd of cattle – a horse or donkey hobbled and tethered to a forked stake, and a small flock of sheep or goats sheltered in a fold of stout branches.

In front of this group of shelters is staked down a long two-stranded leather rope (*daangol*, pl. *daandi*) to which the calves of the household herd are tethered when the herd is in its corral. In front again of the calf-rope is the cattle corral (*waalde*; *baalorgal na'i*). This is often merely a circular patch of earth trampled by the beasts' hooves, for wood is scarce and movement so frequent that a corral fence (also called *hoggo*) is not often made. Round the corral under convenient bushes or trees there may be rough beds of the type found in the ancillary huts. In the centre of the cattle corral is a smudge fire round which the cattle gather in the early morning and in the evening when they return from pasture. The fire is always lit, even in the rain, on the return of the cattle and extinguished after they have gone out to pasture the next morning. The corral fire is known as *duudal*, the root of which signifies 'abundance'; and to make a fire of aromatic wood (*surna*) is a method of maintaining or increasing the herd by magical means.

The importance of the corral fire for the well-being of the herd is attested by a number of Fulani myths.

The whole homestead, consisting of one or more bed-shelters, ancillary shelters, calf-rope and cattle corral, is called *wuro* (pl. *gure*).

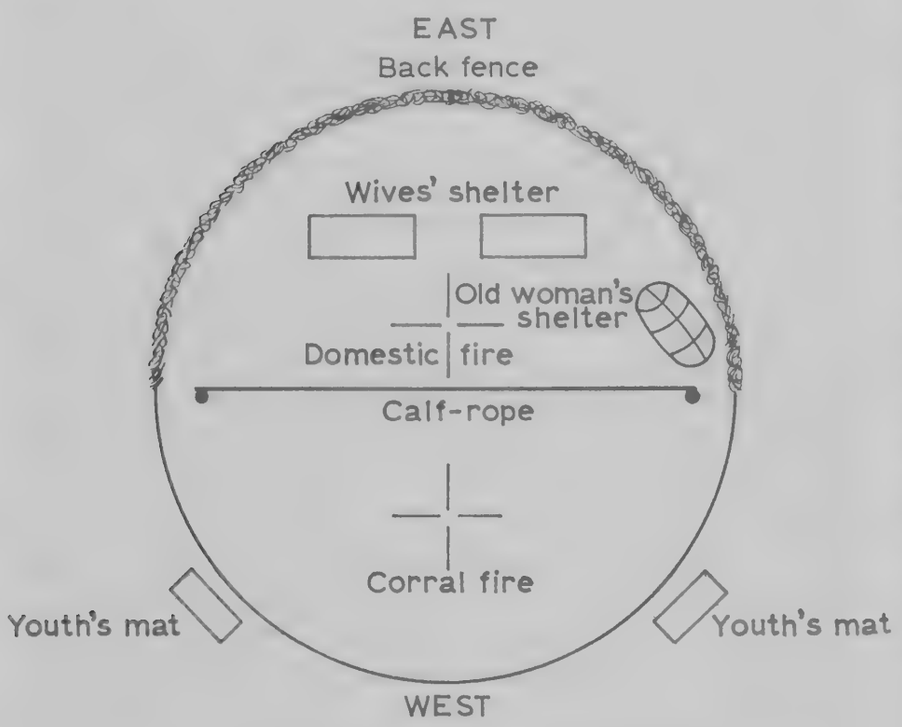


Fig. 9.1 Plan of Woɗaaɓe homestead

As we shall see, camps are formed of a number of homesteads such as this, and the general shape of a camp is a rank of homesteads facing west with their calf-ropes and bed-shelters in the same straight line from north to south. Larger camps are formed by the addition of other such ranks of homesteads pitched to the east and west, but never due east and west, of the first rank. There is no exact term for such a camp; the most general term for a camp is merely the plural of 'homestead' (*gure*).

The bed-shelters with their westward orientation and their north-to-south alignment; the cattle corral in front and, separating them, the long leather calf-rope – these essential features are the distinctive mark of the Woɗaaɓe homestead and symbolise important social relationships.

The whole homestead takes the name of the male household head (*wuro* Adamu, 'Adamu's homestead'). The term is not used in a genealogical sense. The bed-shelters (*suudi*) are the dwellings of the household head's wives, and the matricentral components of a compound family – a woman and her children – are known collectively by this term. *Suudu* may also be used in a genealogical sense; thus *suudu* Raya may mean 'the bed-shelter belonging to Raya' or 'Raya and her children' or 'the descendants of Raya's sons'. The Woɗaaɓe term for a family or household is *iyaalu* (pl. *iyaluuji*); it is never employed in a genealogical sense and is used mainly in formal greetings.

There is a clear division in the homestead between the male and female spheres of action. Between the back fence and the calf-rope is the female area. It is normally the wife's duty to fell suitable trees, or collect them, for her part of the back fence, although she may be helped in this by her sons. The main shelters in this part of the homestead are the *suudi* of the household head's wives, and the essentials of these shelters – the domestic utensils and the bed – are regarded as the wives' own property. The *suudi* are arranged in rank facing west, the senior wife's shelter at the north, the rest in descending order of seniority to the south. A senior wife is described as *wailaajo* ('north-one') and any junior wife as *ɗombinaajo* ('south-one'). The ancillary shelters in this part of the homestead are also to do with women. The *bukkaaru* shelters the aged mother of the household head, the *suura* is the confinement-hut of the household head's daughter in her first pregnancy.

The only males who sleep in this part of the homestead are the household head and his infant sons. The household head has no hut or bed of his own, but sleeps in the *suudi* of his wives in rotation; the wife with whom he sleeps cooks an evening meal for him. Infant boys sleep with their mother when possible, otherwise with their paternal grandparent or with other infants of the same camp in a homestead in which a *bukkaaru* has been built for them.

No males, other than infant boys, eat in this part of the homestead. Although a household head's horse may be tethered there, it is looked after by the infants of the household, who feed it with grass cut by one of the lads of the household and water fetched by

one of the unmarried daughters. If it requires special attention, such as hoof-cutting, bleeding or barbering, it will be led away from the homestead. Apart from the help given by the sons of the household in erecting the back fence and the sheep or goat folds when camp is pitched, or lifting household equipment on to pack animals when camp is struck, males do not work in the *suudi* area. It is not considered seemly for any male of cattle-herding age – that is, above the age of about ten – to go into this area unless asked to do so for some specific purpose by a woman of the household. At dusk a household head hands his infant son or daughter his cap and trousers to place on the shelter of the wife with whom he will be sleeping, for a Boɗaaɗo spends the evening clad only in his long gown. He leaves his wife's bed at sunrise and completes his toilet elsewhere. When going to market, he demands his best gown and slippers and these are brought by his daughter or his wife to a place outside the domestic area. A household head keeps his personal belongings in the shelter of his senior wife. His saddle and horse-trappings stand beside her household utensils on the flat roof of the shelter. What cash he has to carry with him, and documents such as cattle-tax receipts and Koranic scripts, he may hide in one of his wife's prize basketry-covered churns. His best clothes are carefully folded, rolled tightly round a spare bed-pole, covered with clean rags, and lashed with string.

The *suudi* area is given over exclusively to the day-to day execution of women's tasks. A wife maintains all her own equipment. She makes coiled basketry mats for calabash covers and for sale in the market; bark-fibre mats which are used as blankets, roofing, and as part of the harness of the pack-oxen. She scours, sews and decorates calabashes, stains and polishes her bed-poles, lashes and refurbishes the ceremonial items of her household equipment. A wife plans the subsistence of the household. She milks cows and prepares butter and milk for market and for home consumption. She prepares cereal foods which she obtains by purchase or barter, and distributes food to her own children or those of a former wife of her husband, to her husband's mother and father if they are living in the household and, on occasion, to her husband. Infants of both sexes, her own unmarried daughters and daughters of her husband's former wives, and the wives of her sons eat with her at her own hearth, and no female ever eats outside the *suudi* area. A housewife may engage in services for cash, such as spinning raw cotton, and she may attend to the coiffure of certain of her relatives and neighbours. In addition to these tasks she cares for, and gives an elementary education to, her infant sons before they pass into the hands of the male elders. In all these activities she is assisted by her daughters, for whom work can be found at all ages, and who learn the tasks of a Fulani housewife by gradual participation in all of them. Her husband's mother may give valuable help, particularly in caring for infants. Apart from milking and marketing, these day-to-day tasks are carried on within the housewife's own *suudu* area.

West of the calf-rope is the male area of the homestead, where the cattle, the main interest and preoccupation of the men, are kept. Women are allowed into the cattle corral only in special circumstances. A household head's wife, or a daughter deputed by her, may milk her allotted cows at milking times, but not if she is more than about seven months pregnant. When, during the day, a cattle corral is vacated, women do not cross it and children do not play in it. Strangers, particularly butchers and cattle-brokers and those who do not speak Fulfulɗe, are not admitted, and an invitation by a household head to go into his cattle corral is made in ceremonial terms.

Thus only those who have to do directly with the household cattle go into the corral. A household head and his sons spend the early part of the morning, before the women come to milk, around the corral-fire, inspecting, discussing and attending to the household stock. The menfolk of the household eat their meal to the west of the corral. Men eat in order of age – the eldest first – and with an element of secrecy, so that when a man is given the bowl by his senior, he takes it aside behind some convenient bush or tree, or at least crouches with his back to the others, who turn away, affecting not to notice. In large camps, the evening gathering in front of the cattle corral of the senior herd-owner follows this pattern. The senior elders sit nearest the fire, juniors in the next row, and so on; in very large camps, youths are expected to make their own fire elsewhere. When milk is drunk during the evening, both elders and youths drink in pairs of age-mates in order of seniority, retiring to the shadows with their backs to the fire. It is worth noting that the herd around its smudge fire often follows a similar pattern of grading, the lead-bull nearest the fire, the stronger cows near him, young bulls and heifers at the periphery, the calves apart, tethered to their

rope. The identification of the family groups of males with its herd is not uncommonly illustrated by the evening circle being augmented by a ring of the Woɖaaɓe Red Longhorn cattle which forsake their corral fire to join their owners, who say that they have 'come to be admired'.

Only the household head enters the *suudi* to sleep. Boys and youths who are old enough to tend cattle sleep near the corral on rough grass-couches, so that in a large household the cattle may be ringed at night by sleeping herdsmen, who, in the wet season, are at hand to control a herd frightened by prowling hyenas. A newly married man with a childless wife sleeps with her at his post. In the evening she does not leave her husband's mother's hearth until the household head comes to retire; in the morning she gets up early to re-enter the homestead as soon as he has left.

The work of the males is primarily herding, which is usually carried out in some sort of roster operative for a household or in conjunction with another. Although Woɖaaɓe assert that 'the age for herding cattle is never past', a household head falls into an advisory capacity as soon as his sons are strong enough to carry out the roster by themselves. He then occasionally relieves a son who feels he has been hardly dealt with because, for example, his free days have not fallen on a market day. A household head sometimes 'rests' a boy who has had unusually hard herding.

When not herding or absent from the camp on business or pleasure, Woɖaaɓe men sit under the shade of a tree to the west of the cattle corral. Here they receive visitors and exchange news, mend clothes and make cowhide sandals. Their most important job, and one which is continuous, is making the many varieties of ropes and lashings used by them from strips of bark gathered in the bush or from grass fibre bought in the markets.

Women rarely come near the men's shade tree in the daytime, and men rarely go into the homesteads. A man's instructions to his womenfolk are called across the intervening cattle corral, and infants may find employment carrying this and that between the women's shelters and the men's shade tree.

The physical structure of the Woɖaaɓe homestead thus canalises the day-to-day activities of the household and symbolises the relationships of its members, particularly those of different sex.

This characteristic sexual division may be carried yet further. Women approach their own and other households from the rear, men from the front. At death women are buried to the east of the back fence, men to the west of the cattle corral. When moving camp the household rises up and moves in an order of march similar to its day-to-day grouping. Staff in hand, the household head rides to a point west of the cattle corral and leads his herd out westward with the characteristic herding-call of his clan. His sons take up their positions on the flanks of the herd, where they normally sleep when the herd is corralled. Behind the herd comes the small stock, tended by younger children, and behind these again the laden pack-bull of the senior wife, led by herself or her daughter. The bulls of the junior wives follow, mothers and older daughters carrying infants and babies, boys carrying new-born lambs, or urging on sick beasts.

In addition to this institutional evidence, we must note finally a relevant linguistic usage. The root *gor* is unequivocally common to 'man' (*gorko*) and 'west' (*gorgal*, Sokoto dialect); while the root *rew* ('follow') is duplicated in *rewbe* ('women', lit. 'those who follow').

FORMATION OF THE FAMILY AND ITS HERD BY FIRST MARRIAGE

This description of the Woɖaaɓe homestead has necessarily been that of a fully fledged household, with more than one *suudu*, a complement of herd-boys and daughters of the household, sons' wives, a household head's mother and so on. The first stage of the development of the family and its homestead is accomplished in a number of distinct steps embodied in the Fulfulɗe term best translated 'betrothal marriage' (*kooggal*). *Hoowa*, the appropriate verb, has a basic meaning 'to marry a virgin', and the first distinctive feature of this form of marriage is that the partners should be virgin. A second feature is that they should not have been married before. A third feature of *kooggal*, apparent in linguistic usage, is that the marriage is made on behalf of the partners by their elders, usually their paternal uncles and aunts acting for the partners' fathers. Finally, *kooggal* has the implicit connotation of a marriage between partners who are of certain preferred categories of kin.

In describing forms of marriage other than *kooggal*, Woɖaaɓe mention only the immediate ceremony of marriage. In describing *kooggal*, however, they specify a series of events which includes:

preliminaries and betrothal; bride removal and 'sleeping out'; first pregnancy of the bride; name-giving of the first child; and establishment of the homestead.

Preliminaries and betrothal

The preliminary arrangements for betrothal are not marked by any particular ceremony. Informants say that the father of the boy goes to the father of the girl and says: 'Here is my son; give me your daughter, that we may betroth them lest she go off too far.' The girl's father rejoins: 'I give her to you; make a betrothal.'

The sibling group of the boy's father – his paternal uncles and aunts – seek to betroth him at any time after his circumcision (vb. *naada*: *remna*) at the age of seven to ten. It is at this time that the boy's services to the herd begin to be no longer negligible, and his interests in it are now furthered by his father. 'When he has been circumcised, you take your son into the cattle corral, show him the beasts ear-marked for him at his name-giving, and again ear-mark for him what further heifers you are able, so that when he marries he may have their increase.' Animals are not, literally, ear-marked for this purpose, but the term adequately suggests the nature of this event.

When the boy's interests in the herd have been thus enhanced, the future provision of the complementary services of a wife are thought upon – 'There is the herd – the milkmaid is to follow' – and the preliminaries to betrothal are set forward.

Childhood betrothal is common and is in fact almost part of the definition of *kooggal*. In any case, the girl should not be older than the boy, and there is a general preference, confirmed by sample censuses, for an age disparity of about three years.

The boy's representative comes with a witness (*seedowo*) to the camp of the girl's representatives to discuss the matter. This may involve some circumlocution and will be accompanied by particularly 'correct' behaviour, involving lengthy greetings and expressions of mutual respect. The successful completion of the preliminaries is marked, on the next suitable day, by a feast at the camp of the girl's representatives for which a feast-bull is provided by the boy's father. This stage is known by the Woɗaaɓe as *kooggal aranal* ('first betrothal').

When the intended partners are not in their infancy, these preliminaries also include a decision as to the place and time of the next stage, which is the betrothal ceremony proper. Representatives of the partners may become separated by the exigencies of seasonal movement or migration, but this is not a valid reason for the postponement of the betrothal. The connection may be kept up during these years by periodic gifts of single calves given by the boy's representatives to be slaughtered and feasted upon by the girl's people.

The next stage in the procedure is the betrothal ceremony proper (*kooggal*), with which is incorporated to some degree a Muslim ceremony (*teegal juldanku*). This composite ceremony takes place when the intended bride has begun, or is about to begin, menstruation. The burden of both *kooggal* and *teegal juldanku* is borne by the boy's representatives, particularly his father, although the feast is held at the homestead of the girl's representatives. The boy's party come to the camp of the girl's party and say: 'We wish to make a betrothal.' The boy's father provides two bulls to be slaughtered (*kirsaabi*) for the *kooggal* feast. The girl's representatives engage the services of a Muslim scribe, but their expenditure on his personal fee (2*s* to 5*s*) is refunded by the boy's representatives who also pay a similar sum to one of the girl's representatives for his services as her spokesman (cf. Arab. *Wali*) in the Muslim marriage ceremony. Finally, a marriage present (*rubu diina*) is made by the boy's father. The greater part of this (about 8*s*) is given to the bride and is known as the 'head of the wedding' (*hoore teegal*). The remainder, a sum equivalent to his personal fee, is given to the scribe and is known as the 'wedding rope' (*boggol teegal*). The boy's father's final duty is to provide the kola-nuts which are eaten ritually at the Muslim wedding ceremony. Among the Woɗaaɓe the two *kirsaadi* are the important elements in the marriage exchanges at this stage; often the other fees are not included, and in any case no bride wealth, either in cash or kind, changes hands.

Bride-removal and 'sleeping out'

The next stage in the betrothal-marriage ceremonies is the first removal of the bride to live with her husband. If the *kooggal* ceremony has taken place before the onset of menstruation, as is normal among the Woɗaaɓe, the girl goes to live with her husband when his representatives have good grounds for believing that menstruation has begun. The boy's father, so the stereotyped description runs, goes to the girl's father and says: 'My daughter-in-

law has reached marriageable age; give me back my daughter-in-law.' If *kooggal* has taken place about the time of the girl's first menstruation, an interval of up to a year is observed. When *kooggal* has been made early in the wet season, a girl may take up her first residence with her husband at the beginning of the dry season, which is regarded as the best time. If *kooggal* was made later in the wet season, a girl would not be likely to go to her husband until early in the next wet season.

The first bride-removal is called *biiku* or *biikorde*. Although this term is in general use by Pastoral Fulani, its derivation seems to be Hausa: *biki*, a feast, especially a marriage feast, or *biko*, 'trying to reconcile a runaway wife', or 'a joining'. All these ideas are present at this stage of the betrothal-marriage procedure, but that of reconciliation seems to be most noticeable. The attitude of the bride at this period is one of unwillingness to go to her husband's father's homestead. Her expressions of reluctance and displays of shyness may be merely conventional, but they may be sincere, for after the *kooggal* ceremony has been held a girl may secure an annulment of the match by repeated refusals to take up residence. A successful *biiku* thus carries the implication that a girl is 'reconciled' to living with her betrothed husband.

The bride-removal takes place in two stages. The first stage (*biiku aranal*) is marked by the gift from the boy's father to the girl's father of a bull which forms the basis of a feast at the homestead of the girl's representative. There may be an additional gift of gowns in which to array the girl for her homecoming (*wartuki wuro*), for which the auspicious day is a Monday.

The first bride-removal is a significant stage in the betrothal-marriage procedure in that now the boy is no longer represented by his father's siblings – his paternal uncles and aunts – but by his own siblings. Once the arrangements for the feast of *biiku aranal* have been made, it is they who take the feast-bull and the gowns to the homestead of the bride's representative. In the late afternoon of the day of the feast, the bride's father may hand over, through his intermediary, a heifer which is to form the nucleus (*viige fuddirde*) of the wife's own stock. The girl's mother provides her with a few calabashes and spoons; in Bornu these are lashed together in the traditional way with leather thongs and are known as *eletti*.

According to informants, the ceremony of taking away the bride is accompanied by a certain jollity. The bridegroom's party may fill their mouths with milk and try to blow it over her; this part of the feast is known as *poofki kosam*, the 'milk blowing'. They strike her playfully with branches of *barkehi* (*Bauhinia reticulata*), the ritual 'blessing-tree' of the Fulani, or even gĕt non-Fulani guests at the feast to do so. One of the bridegroom's sisters may bring a long wooden pestle, of the type used to pound corn, and threaten to belabour her with it. In this way she is escorted to her husband's father's homestead, where his wife, the bride's *essiiko*, meets her. Striking her gently with leafy branches she takes her into her own domestic area. Informants are unanimous in stating that until a girl has been struck in this way by her husband's mother her *biiku* has not commenced. For two days and nights the girl stays with her; neither her husband nor his father, nor indeed any men, may come near the women's shelters. The father of the boy temporarily relinquishes his right to sleep with his wives and takes a bed at the edge of the cattle compound. At the end of the two days the girl's siblings come to escort her back to her own father's homestead, where she stays for a week, her coiffure now being altered to the style appropriate to a 'reconciled one' (*biiketeedo*).

The bride is then brought back for the *biiku tokkudum* (the next *biiku*) and for the first time she is expected to sleep with her husband. She is not, however, given a married woman's shelter (*suudu*). Among the Woḍaaḅe the husband's bed at the edge of the cattle corral is made more comfortable with the addition of poles and cross-pieces and perhaps a grass roof.

The first period or residence of a young couple joined in betrothal-marriage is known as *waaltordu*; they are themselves called *waaltoobe* (sing. *baltowo* – 'the smooth' or 'surreptitious' sleepers). The unobtrusive nature of this period together is marked by the *biiku tokkudum*, for which there is no feast. The bride's husband's mother gives her food as a member of the household and leads her to her bed. There is nothing in the conduct of the young man on this occasion to draw attention to his special position. He eats with his age-mates as usual and retires as they do after the evening gathering. *Waaltoobe* rise early, the girl taking their bark-fibre blankets into the domestic area of her mother-in-law, with whom she spends most of her time; the boy going immediately into the cattle corral for the morning inspection of the herd. In her daily activities the new wife might appear to the casual observer as

another daughter of the household, and for the *waaltordu* period this is indeed her role. She pounds corn and helps to take care of infants; she helps to handle the calves of her husband's father's herd, but does not milk his cows. She helps to load her husband's mother's pack-oxen and may also market milk for her. The latter, who is said to 'befriend' her, may make her small presents of cash from the market takings, or give her presents of calabashes as a consideration for her help in the household.

First pregnancy

The *waaltordu* period of a bride's first residence in her husband's father's household comes to an end, ideally, when the girl is discovered to be pregnant and goes back to her father or guardian for the birth of her child (*woofa*). Such a young woman is known as *boofiido*.

A *boofiido* goes back to her father's homestead in the seventh month of her pregnancy at the latest. In groups with a high degree of dry-season dispersal this will be arranged at the end of the wet season, if the girl is then in her third or fourth month, lest she should suffer undue hardship by returning in the dry season. She does not normally go before the third or fourth month when her husband and father can 'see her belly'.

Once returned to her father's homestead, the *boofiido* is subject to a number of prohibitions. She no longer sleeps out or with her grandmother, as an adolescent member of the household does. A special shelter (*suura*) is built for her, often of branches of *barkehi*.

The *boofiido* is expected to stay in the *suura* for the period of her pregnancy. During the busy hours – at milking times, and when cattle go off or return from pasture – she keeps strictly to this rule, and inquisitive children may be sharply reprimanded for drawing attention in any way to her presence. During the later part of the morning or in the afternoon she comes out to wash herself or have her hair done in a simple, unornamented style appropriate to her condition. At this time of day, when few people are likely to be about, she suns herself or takes the air in the shade of a tree, according to the season. She even helps in a desultory way with household tasks and plays with the young children. In the evening she does not sit near the fire in the women's gatherings lest one of the men should see her, but she is likely to come out of her hut at that time. She is not supposed to draw water, milk cows or sell milk or go to markets and feasts. She is restricted in her use of cosmetics, such as henna and antimony, and jewellery. She wears a plain headcloth and her gown is of white homespun. When camp is moved she follows well in the rear attended by a younger sister or one of her mother's co-wives.

These prohibitions have the effect of excluding the *boofiido* from contact with menfolk or with cattle. Her husband has no contact with her during this period and is not expected to go and visit her. He also avoids her mother and father, while the *boofiido* must herself avoid all menfolk, even those of her own family.

Naming the first child

Shortly after the first child is born it undergoes a naming ceremony (*innde*) which, in essentials, is similar to the Muslim *'akika*, the recommendable naming feast. The canon law recommends that at this feast the child should be named, its head should be shaved, an animal should be sacrificed, the meat eaten within the family circle and the greater part of it given away in alms to the poor. The blood of the sacrifice should not be smeared on the head of the new-born, nor should this occasion be used for the circumcision or excision of the child. If the offering of the *'akika* has been neglected on the seventh day it can be done afterwards, even by the child itself when adult.

Both in their description and their practice of the *innde* feast Woɖaaɓe show themselves aware of these provisions and describe the ceremony as an Islamic one. 'The child should sleep for a week, we should make a name-feast, we should sacrifice – thus the Book says.'

The importance attached to the *innde* ceremony, particularly for first-born children, is seen in strict compliance with the injunction to hold the ceremony seven days after the birth of the child. But the lapse of seven days is not always adhered to, owing to the Woɖaaɓe belief in the inauspicious nature or otherwise of certain days. For example, an *innde* falling on a Wednesday or Saturday is regarded as inauspicious for the child and its future health; if it falls on a Tuesday it portends good luck to the child and its family. Consequently Woɖaaɓe make a point of postponing an *innde* which should have fallen on a Saturday, or bringing forward one which should have fallen on a Wednesday. Nevertheless, an adult name-giving is

unheard of in Pastoral Fulani society, although approved by Muslim canon law. A child is not regarded as a person until it has been named at such a ceremony; a child in its first week is called 'it' and no distinction is made between a stillborn child and one which died before being named.

The head of a child is shaved at its *innde* ceremony, a boy's head being shaved by one of the adult men of the camp in which he is born, preferably his father's brother or, in the case of a first-born, his mother's brother. The 'infant's hair-style' given at this time to ward off disease or misfortune is not changed till a boy is circumcised. Some clans use the naming ceremony to tattoo the clan marks on the infant's face.

Woɗaaɓe do not adhere strictly to the use of a ram or goat at this ceremony; they prefer the feast animal to be one from the herd of the father or paternal grandfather of the child. In the case of a first-born the maternal grandfather also gives a beast, the two name-giving beasts being known as *bohidi* (lit. 'first-child bulls'). The name-giving of a first-born in particular should be marked by the slaughter of bulls in preference to rams or goats, and the feast animals should not have been bought.

A child's *innde* ceremony is normally held at its father's homestead, but that of a first-born is held at the homestead of its maternal grandfather. In the case of a first child (which concerns us particularly in this instance) both the maternal and paternal grandparents provide feast meat, and it is common to invite to the feast as large a circle of kinsmen and neighbours as possible. Neither the father nor the mother and child are present. The father stays in his father's homestead and takes his turn of duty with the cattle on that day so that 'his brothers may go'. 'He is herding' is often used euphemistically to refer to avoidances of this sort. On the morning of the feast the mother stays with her child in the *suura*, where she receives kinswomen and friends who come to bring their contributions to the feast and to see the child. When, in the middle afternoon, the naming ceremony and the men's feasting become imminent, she goes off with another woman, taking the child, to a nearby homestead, or even a quiet place in the surrounding bush, where she spends the rest of the day.

The name-giving is performed by the menfolk and the Muslim scribe who is called in to officiate at the expense of the child's father or grandfather. The proceedings begin with a distribution of kola-nuts by the child's father's representative; these are to be eaten ritually after the prayers which now take place. The scribe then inquires if the beast has been slain according to law, even if he himself performed the killing earlier in the day. The representative of the child's father then asks what the child shall be named. The scribe counters by asking its sex and on what day it was born. On being told he pronounces a name appropriate to that day and enjoins them to call the child by this name. Each day of the week has at least one masculine and one feminine personal name appropriate to it, and these names are invariably renderings into Fulfulɗe of Arabic personal names which have become current in various forms throughout the Islamic world. These are known to the Woɗaaɓe as 'religious names' (*innde Mohammadiiya*) or 'scribe's names' (*innde mooddibbe*); they are to be contrasted with names received or taken later in the life of the child which are the private names of the child, its 'grandparents' name' (*innde maama'en*), its nickname (*innde soomoore*) and an adopted name (*innde jammoore*).

Making the homestead

As we have seen, the *boofiido* stays in her father's homestead for a period of two to two and a half years, expressed by Woɗaaɓe as 'three wet seasons' (*duubi tati*). Her return to her husband takes place at the end of this period and is known as *baŋtal*. The root *baŋ* suggests 'superiority', 'dignity', 'coming back to a more desirable state'.

It is on this occasion that the parties to the betrothal marriage join in providing the partners with the symbols of full marital status which constitute a homestead. The husband is now said to 'know his cattle' and is, in effect, a herd-owner. The wife is formally elevated to the status of wife and mother, and acquires the right to milk in her husband's herd.

The obligation of the husband's party on this occasion is the 'home-building' (*ŋibol suudu*). The husband's brothers clear a homestead space and erect a back fence. The husband's father makes a calf-rope and its stakes and this is set in front of the homestead by the husband's mother. The husband's father's wives make a number of mats from bark collected by their sons. These, and a rope harness made by the husband's brothers, constitute the trappings of the pack-ox (*dimdiri*) furnished by the husband's father. Woɗaaɓe say, 'without a pack-ox there is no home'. When

these items have been assembled, the husband's father buys a number of women's gowns and head-cloths, and his sons take these with the harnessed bull to the homestead of the bride's father, where a feast for which a bull has been slaughtered awaits them. A conventional period of one week elapses before the complementary part of the *baŋtal*, the 'collection of the calabashes' (*moobol kaakul*), takes place. The wife is arrayed in the gowns given her by her husband's father, but her child is clothed by its maternal grandmother, who may deck an infant boy with a bead apron, cowhide bracelets and a collar embroidered with cowrie shells which bears the same name as the halter which attaches calves to the calf-rope (*daande*). This infant's regalia will be kept by the wife until the first occasion on which she sends a *koowaado* daughter back to her husband with child. The bride is provided with a bed by her mother. The rugs and mats are made or bought by herself and her co-wives. The bed-poles have been cut and trimmed by the bride's brothers, and stained with a red dye made from *genki* wood by her sisters. The tokens of a dairywoman's occupation were given to the bride by her mother at her *biiku*, in the form of the *eletti* (see p. 191). These are now supplemented by a formal domestic portion of the same design, known as *kaakul*, and consisting of calabashes, spoons, churns and dippers lashed together in a traditional way with leather thonging and basket-work. A *kaakul* attached to a dismantled bed ready for loading on to a pack-ox is known as *shinda* (cf. 'to be split').

The wife's pack-ox is now loaded with these gifts, and the bridal party sets off for the husband's father's homestead, where a further feast, for which a bull has been slaughtered, awaits them and lasts as long as the resources of the donors permit, usually three days. The *kaakul* and bed are set up in front of the *suudu*, and the presents of the bride's party are displayed.

The husband is present at the feast but no undue prominence is given to him in the celebrations. He sleeps with his age-mates during the period of the feast. His father does not sleep in the homestead but has a *suura*-like shelter erected for him west of his own cattle corral. The new housewife sleeps in her husband's mother's shelter during the feast, and the couple enter into their roles of householders only on the first propitious day after the feast, which may occur after camp has been moved.

The new housewife now attaches to her homestead *daangul* any calves belonging to her husband's cattle or her own cattle, if she has any, and also the calves which may have been given by his maternal and paternal grandfathers to the first-born child. Fully grown beasts of her husband's and her own herds stay for the time being in the paternal grandfather's herd, being divided only when the latter's youngest son has married and made his own homestead.

THE MARITAL CONTRACT

When the bride finally takes up residence with her husband, this is a complete recognition of her status as wife, mother and dairywoman. The contributions of the two parties are now brought together as the essential items of a Pastoral Fulani homestead, and at the same time symbolise the complementary relationship of husband and wife in their family life.

As we have seen, the bride's gifts are prepared and given to her by her own family. Her bed is the distinctive mark of a married woman's hut, for it is found in no other shelter. When a woman dies her bed is left in her shelter and, after a period of mourning, the camp is vacated. To take a dead woman's bed is an unthinkable, not to say impossible, act, for deserted homesteads (*wiinde*, pl. *wiinde*) still recognisable as such are never entered lest dire misfortune should befall the trespasser. The wife's traditional household portion, given her by her mother, is a symbol of her reproductive powers. Before she has first conceived it is a miniature. When she has borne her first child it is commensurate with her status as a young mother (*kaabo debbo*). Thereafter she adds to her stock of working calabashes as her family grows and her responsibilities for feeding children increase. When a girl-child is born to her she adds decorated calabashes to the ceremonial portion. The *kaakul* is never in everyday use but is kept over a wife's bed, to be refurbished when auspicious days occur after camp has been pitched, and displayed proudly at camp feasts. As her daughters marry, items of the *kaakul* are gradually given away until, as an old woman with no milking rights and no responsibilities for feeding children, there remain to her only one or two undecorated calabashes. Finally, her mother may give the bride a female calf from the maternal herd, if one is available.

One of these three gifts, the bed, perishes with its owner; the other two are inherited in the maternal line (*lennyol dewol*) and,

indeed, appear to be ritually devoted to this specific purpose. The wife's family furnish her in this way with objects which are both essential to, and symbols of, her domestic and economic services and her reproductive powers.

While she is married to him, the husband's party make available to her the transportation which is essential to a nomadic way of life, and is also the symbol of her relationship with the herd which comes into being as the result of her marriage. She cannot inherit these goods and they are, in a sense, loans. The pack-bull upon which her goods are loaded when camp is moved is forfeit if she runs away; but if her husband dies she may be allowed to use it until she is provided with another by a new husband. The first pack-bull having been provided by his father, the husband must thereafter see that his wives have sufficient animals, by training calves to the work or by borrowing from neighbours in an emergency. Finally, the calf-rope is laid down as an essential part of the homestead.

The symmetry seen in the layout of the homestead and the apportioning of household tasks is amplified by this consideration of the objects presented to the bride and the husband in the course of the *baŋtal* feast. For the objects inherited by the wife from her mother are precisely those which are essential to that part of the homestead which is the woman's sphere of domestic activity. She has no rights of ownership in any object made available to her by her husband's party, or in the patrimony which he receives on this occasion.

Similarly, a man receives simply his patrimony and has no right to expect, say, a dowry of cattle from his wife's party. He has no obligation to provide her with more than a suitable domestic site. The crucial symbol in this complex is the calf-tether, which at once separates and connects the male and female elements. It is a rope, and therefore is made by men; it is connected with cattle and is therefore provided by the cattle-owning group of which the husband is a member. Of all the many kinds of rope used by Pastoral Fulani it is the only one made of cowhide. It is provided not merely because a man has married – it is first staked out when his senior, betrothed wife returns with her first child. Although it is her duty to see that it is taken up on striking camp and restaked when camp is pitched, it does not become the wife's property, and is regarded as being put up 'because of her'. The phrase 'my calf-rope' is never used by husband or wife; it is always called 'the calf-rope' or 'the homestead calf-rope', and in an analytical way a Fulani may describe a certain camp as 'so many calf-ropes'. As we have seen, the calf-rope dividing the corral from the domestic hearths also divides men and women in their everyday life. But it is at the calf-rope that the only routine co-operative work of man and wife is done. When the herd does not require examination or treatment, husbands frequently join their wives at milking time to help release the calves from the rope and return them to it when they have been suckled. The day-to-day management of the calves is not the direct responsibility of either the husband or the wife. No herdsman will herd calves and, when they are released to pasture about the camp, this is the job of the small boys and girls under what appears to be the casual supervision of either men or women. If a bunch of calves gets to the herd and sucks so that the household is deprived of a milking, the blame is laid squarely on the children. 'Children play about the camp all day with the calves, and they should keep an eye on them.' The term for a group of age-mates in a dance is 'a calf-rope'. Just as the *kaakul* expresses the size of a woman's family and thus her reproductive capacity, so the calf-rope expresses the size of the householder's herd and its reproductivity. The only time a man crosses the calf-rope is when he lies with his wife, and the only time a wife crosses the calf-rope into the corral is in order to obtain milk. When a man has a first wife and a child he may hold his herd; when a *koowaado* has borne her first child she may assume milking rights in her husband's herd. These rights are assumed at the *baŋtal* ceremony of home-building, the symbol of which is the calf-rope. The marital contract thus appears to be one in which calves, the increase of the herd, are identified with children, the increase of the family. The marital duty of a husband is to keep the herd at such a size as to be able to provide for his dependants; the wife's duty is to bear children who become the labour force of the family. A wife expects from her husband milk; a husband expects from his wife children. The whole stability of Woḍaaƀe marriage, and the economic efficiency of the family of which it is the origin, is bound up with the fulfilment of these expectations.

THE VALUES OF MARRIAGE BY BETROTHAL

As we have seen, *kooggal* lies at the heart of the traditional values of Woɗaaɓe, and indeed lay at the heart of their social organisation at a period when the clan had a more markedly corporate organisation than it has today. Although altered and diluted by their increased participation in the wider society of the Western Sudan, brought about by the cessation of warfare, by population increases, the spread of the Muslim religion and their closer incorporation into the Islamic State, these traditional elements of Pastoral Fulani social organisation are continually borne in upon the field observer. It is important to note that, although the elements of social control exercised by the *maudo laawol pulaaku* have diminished considerably, the marriage system with which that control was associated is still described as the most important of the traditional institutions. It appears, even today, to be capable of affirming and indeed effecting the 'differentness' in which the Woɗaaɓe believe their chief virtues to lie. If the endogamy of the Woɗaaɓe in the past was total, it was the endogamy of the alien pastoralist in a strange land. The partial and graduated endogamy of the Woɗaaɓe today is more akin to the endogamy of the caste.

The importance given to *kooggal* in verbal statements is corroborated to some extent by its actual implementation. In the table below, the validity of the relation we have suggested between *kooggal* and the establishment of a homestead, a household and a herd is set out. The successive completion of the three main stages in the sequence – betrothal (*kooggal*), first residence (*biiku*) and constituted homestead (*baŋtal*) – are enumerated with reference to a sample of Woɗaaɓe male householders:

BETROTHAL-MARRIAGE AND THE ESTABLISHMENT OF HOMESTEADS

Total number of male householders in sample	39	100%
Group A. Number achieving all three stages (betrothal, first residence and establishment of homestead)	28	72%
Group B. Number achieving first and second stages (betrothal and first residence)	5	13%
Group C. Number achieving first stage only (betrothal)	2	5%
Group D. Number never betrothed	4	10%

Ninety per cent of the male householders in the sample had thus commenced their marital histories in the traditional way, while only 10 per cent, for one reason or another, had not. These figures suggest at least a nearly universal intention on the part of Woɗaaɓe to betroth their children. Evidence from genealogies confirms this high betrothal rate. However, the figure of 72 per cent actually completing all the requisite stages shows that this intention is not effectively realised. An examination of the circumstances surrounding the failure to achieve each successive stage is required, for clearly such an important means of attaining marital status is not lightly abandoned. Since betrothal-marriage concerns not merely the marriage partners, but also the parties who initiate it and carry it out, the availability of competent persons in this category at every stage is necessary to its success.

In the sample there were four householders who had never been betrothed (group D). The father of the first of these had lost all his cattle in an epidemic and, from an early age, the informant had been sent to the household of his mother's brother for whom he acted as herdsman. In the second case the informant's father died in his infancy and he was brought up in the household of a classificatory mother's brother. The third of these informants, a man in about his fortieth year, was chronically leprous and half-blind, and was dependent upon his kinsmen for assistance in herding. He had gone through a form of marriage with a girl from a different clan when he was thirty, but this had not lasted long. The final case was one in which information could not be elicited. Since a younger brother was later found to have established a homestead with a betrothed wife, it is at least likely that the informant had in fact been betrothed; he belongs to this category by default.

In the sample there were two cases where informants had reached the betrothal stage only, and the arrangements had lapsed before the *biiku* ceremony (group C). The father of one of these had died when he was a youth. He had lived with his father's brother who took advantage of this opportunity to split his herd, sending the youth to attractive but distant pastures, where he spent several wet seasons. He contracted a form of marriage with a girl from one of the clans whose home pastures were in the area in which he stayed. The *koowaaɗo* of the second informant died before their *biiku* ceremony could be performed, and he was allowed to marry an immigrant girl from a strange clan.

Group B concerns those betrothal-marriages which fail during the first residence of the bride. In our sample there were five such cases. In two of these the wife was childless for longer than was

expected by her husband and his father. In one of these cases the husband divorced his wife with the support of his father. In the other case the husband introduced another wife into the household and, when the latter bore a child, the childless *koowaaɖo* ran away. In the third case the boy's father died during the period of first residence of the young couple. The husband's guardians were his elder brother and his mother. The wife ran away from a domestic situation in which the two men were dominated by their mother. In the fourth case the wife ran away to a lover in another clan. In the last case there was a quarrel between the two full brothers who were the fathers of the marriage partners, and the girl's father had the betrothal annulled.

Finally, we must review the conditions in which the establishment of a homestead by betrothal-marriage was successfully achieved (group A). In sixteen of the twenty-eight cases in our sample, the boy's father was alive and present at all stages of the arrangements. In three cases the father died during the first residence of the bride; in two of these the girl lived with the husband and his mother, in one with her husband and his elder brother. In seven cases the father died after betrothal but before the son's wife took up residence. The husband's guardian was in one case his father's brother, in four cases his elder brother, in one his mother and in the other his sister's husband. In one case the father of the boy made the preliminary arrangements and then departed on a pilgrimage to Mecca. The subsequent stages were carried out by his younger brother. There was one case in which the father was not alive at any stage. Here the custody of the boy was given to his mother, whose father betrothed him within his own clan, but died between the preliminary arrangements and the first residence. His son, the boy's mother's brother, carried out the remaining stages. The true filiation of the boy was not forgotten and he eventually offered allegiance to a minor chief of his father's clan.

These examples illustrate a number of general characteristics of betrothal-marriage up to the stage where the homestead is set up and the herd constituted. The first and most obvious of these is that the partners in *kooggal* shall not too flagrantly defy those physical characteristics which this form of marriage is held to maintain. If the congenitally infirm or deformed survive the rigours of a childhood in Pastoral Fulani camps, they are not betrothed since their condition carries with it a sort of moral stigma. This is particularly true of those who are in some degree congenitally insane. They are believed to have been 'caught by the wind' (*hendu nangi be*) and, like the sand-laden whirlwinds, they wander about the bush. Social conventions are not applicable to them, they are not betrothed or married or allotted cattle. As we have seen, in the days of *pulaaku* the converse was also true.

As Woɖaaɓe say: 'He who flouted social conventions was himself seized by them. His clansmen, his "body", drove him out and made him live alone in the bush. He went unkempt like a madman.' In field-work in West Bornu, no insane person of this sort was encountered by report or otherwise, but in another area in Northern Nigeria one such outcast was met with. He was under the supervision of his lineage head, who sent him food at his shelter in the bush.

The clan-endogamous nature of the *kooggal* arrangements presupposes a certain stability in the balance of the clan segments which in-marry and inter-marry. An important cause of instability lies in the vagaries of the pastoral life. A number of recorded cases, from this sample and others, illustrate at a personal level the way in which sudden losses of cattle, urgent pressure to migrate or the constant demands of good husbandry undermine that stability by preventing the marriage arrangements which are its best safeguard.

Physical incapacity and pastoral contingencies appear to operate mainly in preventing *kooggal* arrangements from being initiated at all, or in the stages before the *biiku* ceremony. Until the *biiku* ceremony, also, the partners are under the jural control of their parents or guardians, and the dissolution of a betrothal before this stage is a matter for them. If the boy's representatives wish to dissolve the betrothal, they forfeit their expenditure on the bull of the first betrothal and any intermediate gifts of bull-calves. Where the girl's party desires dissolution they are liable for the return of these prestations.

But after the *kooggal* ceremony, which takes place when the partners are 'fit for marriage', they are no longer the pawns of the elders and have the right to appeal for dissolution, or to adopt other courses open to them which will free them of their obligations.

A boy who openly rejects his *koowaado* at this stage would be regarded as foolhardy; he would run full against his father's authority and jeopardise his proper interest in the establishment of

his own homestead with its household and herd. Where his marriage is with a close cousin, he is likely to put himself out of favour with the elders of his own and neighbouring camps. The foundation of a man's status and economic future is laid at his betrothal and depends upon his assumption of the cattle which have been set aside for him. We should make the point that the extent to which the Woɗaaɓe avail themselves of alternative means of gaining a livelihood is negligible. In ninety-eight pastoral households examined in West Bornu in the course of sample census work, only one male had moved out of West Bornu to seek employment for wages. Some orphans are taken into the households of chiefs who are village heads, where they become retainers and do not own herds, but in due time they are betrothed by the chief to daughters of other village heads or prominent cattle-owners, and thus promote by marriage ties the stability of a village head's following. *Kooggal* for Woɗaaɓe men is still the right, and the most profitable, way to start out in the world.

Similar factors operate to deter a girl from seeking the dissolution of her betrothal, and the attractions of achieving a properly constituted *suudu* are powerful. In the past, the sanctions of *pulaaku* operated to keep girls within the marriage system of which their lineage formed a part. But today, unlike the young men who are still tied to property inherited patrilineally, girls have another course open to them if they wish to avoid a match which is distasteful – they may run away to a man outside the jurisdiction of their father's chief. The effect of these unions will be described below.

Dissolution of a betrothal-marriage in the *waaltordu* period when the girl is no longer a presumptive virgin, is similar to dissolution before the *biiku* ceremony, but with certain differences. The formal length of the *waaltordu* period is about one year, and it is clear that the expectations of the household into which she has married, and particularly the expectations of her husband's father, centre upon an early conception. If this does not occur in due time, the marriage exhibits signs of instability. A father may reproach his son by what amounts to an accusation of impotence, and this is the only occasion on which impotence may be openly imputed to a Fulani man. The *koowaaɗo* has to work for her husband's mother, and where the latter has no daughters or is an only wife, this may prove arduous. Although married by Muslim law and Woɗaaɓe custom and expected to cohabit as man and wife (by which terms they are described in this period), the young couple are in all respects treated as single. They are expected to join in the activities of youths and maidens and play a leading part in dances. But the courtship which is the main characteristic of these activities is denied them. A young man who is not enamoured of his *koowaaɗo* may take a more congenial second wife; a girl may refuse sexual intercourse to an unwelcome husband. The *waaltordu* period bristles with problems of sexual, emotional and familial readjustment, which the Woɗaaɓe organisation of the matter seems hardly designed to resolve. When the young bride's conduct as a member of her husband's father's household is above reproach, or if there are few daughters available, but many sons and a big herd, the home-making ceremonies may be carried out in a somewhat summary fashion after a year or so even if the girl has not conceived. What is more likely, and is especially in evidence among certain sections of the West Bornu Woɗaaɓe, is that the difficulties of this period will drive the girl away. She may run off to her father or guardian; if nothing can be done to reconcile her to her new life, the marriage is annulled by agreement of the contracting parties with the consent of their chief, and the marriage-bulls will not be reclaimed. But a young woman at this stage is more likely to run away from her parents' jurisdiction than face their recriminations.

Another cause of the disruption of *kooggal* arrangements is the death of either of the partners at any stage. This is regarded by the parties to the marriage as their common misfortune, the 'work of Allah', and no refund of the marriage cattle would be considered by either party, who both mourn the dead *koowaaɗo*.

A final cause of failure at all stages of betrothal-marriage is a deficiency in the composition of the sibling groups which arrange them. The most important of these is the father of the boy, who nominally initiates each stage of the betrothal with the active help and representation of his siblings. The father's role is most important in the earlier stages of the arrangements; once a *kooggal* tie between the families has been set up by the acceptance of the boy's father's prestations, the arrangements are more easily continued by his successors in the event of his death or absence. In the case of patrilateral parallel cousins, actual patrilateral and matrilateral cross-cousins, there are a number of affines and agnates of the dead man immediately involved in the match, some of them as members

of a marriage party. But where these kinsfolk are not present or where the parties to the marriage are genealogically more distant, then the duty of continuing the arrangements devolves more directly upon the father's successors who are, usually, his younger brother or his eldest son.

The illustrative cases suggest what is confirmed by informants, that an elder brother is not competent to continue a betrothal arrangement, particularly a genealogically close one, since he is of the same generation as the boy, whatever the difference in their ages. A mitigating case is where the widowed mother of the two boys is able, through her kinship ties, to carry out a betrothal. More reliable in carrying out the dead man's wishes and obligations is his own brother. It is he who inherits the widow of the dead man and, because of his status as guardian of his brother's unmarried children, may arrange or continue their betrothals. He also inherits, as trustee for his unmarried children, the dead man's herd.

In *kooggal* marriage we see a traditional institution which is intimately linked to the inheritance of cattle, the establishment of herds and the status of both men and women as married adults. Although Muslim elements have been incorporated into its rituals, it continues to occupy an important place in the wet-season celebrations. It is binding upon men, because of its connection with the dominant form of property in Woɖaaɓe society. Women are freer to refuse its provisions. This is due partly to the greater opportunities for social intercourse between members of different clans, and partly to the change in the organs of social control among the Woɖaaɓe.

Notes

1. A sub-group of the pastoral Fulani people (Ed.).
2. Islamic 'holy war' (in this extract refers to the nineteenth-century *Jihad* in the Western Sudan) (Ed.).
3. A single member of the Woɖaaɓe group; also used as an adjective (Ed.).
4. Language spoken by the Fulani (Ed.).

THE BRITISH FAMILY TODAY

H. Gavron

To complete this summary of changes affecting the family in this country, we must consider the family today. In demographic terms, it can be said that the average marriage produces two to three children as compared to numbers in excess of six one hundred and fifty years ago. On the other hand the expectation of life of its members is now at least seventy, which means that every family has to face the problem of how to care for the very old. It is no longer dominated (demographically) by women, as the reduced infant mortality rates ensure that the majority of male offspring survive. The increased popularity of marriage at ever earlier ages means that the conjugal family spends the greater proportion of its time as a unit of two. The family is also supported in a great many ways by the State today, which was not true in the nineteenth century. Health, education, and to some degree housing, are the concern of the State. Compulsory insurance provides some protection against misfortune, and the tax system is geared to favour the family as against the single. It should not, however, be thought that this support to the family has been at great cost to the nation. There is no evidence to support the view that 'the admitted and always approved benefits of greater material security have been purchased at the price of a steady weakening in the structure of society'. This support is a necessary concomitant of an industrial society which both reduces the power of the family to be a self-contained unit, and demands of it extremely high standards of care for its members. As McGregor[1] suggests, 'Nineteenth-century industrialism created an urban society in which only affluent families could self-helpfully discharge their functions and responsibilities.'

Some critics of the modern family have argued that these 'functions and responsibilities' have been considerably reduced in the period since industrialisation. The modern family, say MacIver and Page,[2] has been stripped of all non-essential functions and its role today is simply to harmonise three major functions, namely the satisfaction of sexual needs, the procreation and rearing of children and the provision of a home. Talcott Parsons[3] suggests that in a 'highly differentiated society' the functions of the family 'are not to be interpreted as functions directly on behalf of society, but on behalf of personality'. The two basic functions of the family are therefore: (1) the primary socialisation of the children, and (2) the stabilisation of the adult personalities of the population of the society. This accords with the view of Bryan Wilson,[4] who argued that the family has simply become a 'highly

specialised agency for affection'. However, to present the family's functions in so limited a way is both to misunderstand the past and misrepresent the present. As Nels Anderson[5] suggests, 'the idea seems to be current that if the family changes it can only be for the worse, and that such change endangers society'. Indeed the family has changed as has society, in the past one hundred and fifty years, but the general direction of that change has been towards increasing complexity of function rather than the reverse. Thus, as Ronald Fletcher[6] says, 'the family is now concerned with a more detailed and refined satisfaction of needs than hitherto, and it is also more intimately and responsibly bound up with the wider and more complicated network of social institutions in the modern state than it was prior to industrialism'. As far as the essential functions described by MacIver are concerned – sexual, procreative and the provision of a home – higher standards are demanded by society today than ever they were in the nineteenth century. Parenthood and the care of children is now a highly self-conscious affair in which the maintenance of a high standard is insisted upon, and the pitfalls are for ever being exposed. The importance of the home today is underlined continually by the mass media particularly through advertisements, the house in particular is usually shown as being well stocked with appliances and comforts, clean, bright and up to date. The revolution of rising expectations ensures that the desire to improve one's home is one of the constant preoccupations of the family. As far as the satisfaction of the sexual needs is concerned, higher standards are asked of the family than ever before, particularly as fertility is now susceptible to control. The concept of romantic love drawing its inspiration largely from legends about knights and ladies is still with us, and is reinforced by the pop song, the advertisement, the magazine, which never cease to extol the virtues of love as a solution to all problems and as a basis for all thought. In a society in which divorce is tolerated if not approved, marriages are expected to provide positive advantages to the partners in order that they continue to stay together. If the family's essential functions are possibly more complex than in previous times, it has also acquired more functions. Economically the family functions as a unit of consumption, as a provider and improver of homes, as a possessor and maintainer of property. As far as education is concerned the extension of State and private provisions has by no means served to reduce the functions of the family in this sphere. The demands of a complex industrial society are such that education is the keystone to adult success. The family must prepare the child in its early years in such a way that it is ready to receive an education and be receptive to it. Throughout his attendance at school the child will need support, assistance and direction from his parents, particularly when he nears the end, and a multiplicity of choices are available upon which his future may depend. Just how important this can be has been revealed by recent studies of the working class and education, which demonstrate to what extent working-class children are handicapped by their family's lack of understanding of the education system, and their inability or unwillingness to support their children emotionally and intellectually throughout their school career. As is the case with publicly provided education, Fletcher suggests the wider health services have by no means taken over all the functions which previously were performed by the family. The family is still the prime guardian of its members' health, and the standards demanded of it are now incomparably greater than they were even fifty years ago. To these functions, now extended, must be added recreation, something that was until recently purely a province of the upper classes. The 'weekend' is now a time of leisure and family activity which is only partially provided for by television and the growth of the entertainment industry.

In fact, says Fletcher, the main functions of the family in our society can be summarised as follows:

1. It provides a way of regulating sexual behaviour.
2. It gives a legitimate basis for the procreation and rearing of children.
3. It provides sustenance and care for its dependent members.
4. It is of primary importance as an agency of socialisation and of education, and thus of the transmission of culture.
5. It bestows titles, roles and duties on its members which are recognised and applied by society (son, husband, wife, etc.).

The next point to consider in any analysis of the modern family is the question – is it inherently unstable? Once again there are many people, both sociologists and laymen, who are convinced that it is. Sociologists such as Zimmerman[7] or MacIver, who have seen the history of the family as a succession of types, have all

tended to regard the modern family in rather a pessimistic light. The detailed arguments of those who feel the family today to be unstable, with specific reference to Great Britain, rest largely on the following assumptions:

1. The family today is an isolated unit unable and unwilling to seek assistance in its distress from wider circles of kin (as it has always done in the past).
2. There has been a loss of authority – primarily moral – within the family, giving rise to ever increasing delinquency.
3. Sexual morality is at a low ebb and is continuing to decline, as evidenced in the increase in venereal disease, illegitimacy and divorce.
4. That family life has been seriously endangered by the increase in married women working who thereby neglect and abandon their children and husbands, creating a host of problems as a result.
5. That its role in society has been undermined by the Welfare State.

As has been demonstrated above, the functions of the family have been extended in this century rather than curtailed. Basically, many of those who argue that the contemporary family is unstable have in their minds the Victorian middle-class family as an ideal norm from which we have strayed. This is conceived of as a perfect example of authority, stability and moral rectitude which provided a linch-pin for a virtuous, stable and moral society. Compared with this ideal, the modern family has suffered a loss of parental authority. This, however, is of course extremely difficult to analyse because in such matters there is always a gap between ideal and reality. Certainly the Victorian father was accorded more authority by society than his equivalent today, but there are few who would today argue in favour of the complete dictatorship of such a father, and a great deal of the legislation in the last one hundred and fifty years has been directed quite deliberately at reducing this authority and giving equal status to the women and children. In a more subtle form, however, it could be argued that there has been not so much a loss of authority as a loss of contact between parents and their adolescent children, which may well be *one* of the factors contributing to the undoubted increase in delinquency in all such societies as our own. But the blame cannot be laid specifically at the family's door; the problem of delinquency involves the whole of the society we live in, its material and psychological attributes, its ideals and its ability to live up to them. Thus while it might be fair to argue that rising delinquency rates are signs of strain within the family, they are not conclusive evidence of its increasing instability. Much the same arguments can be advanced on the subject of sexual morality. Increasing divorce rates must be seen against increasing marriage rates, and a falling average age of marriage. Two world wars this century have produced great social confusion from which we have slowly recovered. What is more, as Rowntree and Carrier point out, divorce rates are to some extent a product of the legal provisions in existence. It is also worth noting that Rowntree and Carrier[8] found that marriages contracted since 1945 show some signs of being more stable than those contracted in the previous twenty years. Increasing illegitimacy rates may in part reflect the disorientation of the young, though as previously mentioned this is not totally the fault of an unstable family; it also reflects the ambivalence in our society towards sexual behaviour, with a widening degree of permissiveness in private, while in public the Victorian attitudes remain. It also may be a result of the rise in status of women, who no longer feel impelled to treasure their virginity and thus feel freer to enter into premarital sexual relationships.

The controversy over women working [is worthy of closer examination.] As far as the stability of the family is concerned there is little indication of serious neglect, and it is certain that many *working*-class children get far *more* parental care today than they did one hundred years ago. The typical working wife is in fact not the mother of young children, but is likely to be aged between forty and fifty with most if not all of her children grown up. Studies such as those by Yudkin and Holme[9] reveal no evidence of severe damage to children, and these studies revealed, as did Viola Klein's,[10] that the majority of working wives have the support and approval of their husbands. Once again there is little evidence here to support the view that the family is unstable and in decline.

What is true is that the family today gives much more recognition to the individual as a self-sufficient independent person. Increased mobility and the decline of hereditary privilege have

combined to reduce the power of the family to confer status and increase the responsibility of the individual to find at least in part his own level, through achievement rather than ascription. Family relationships today may be less dependent on outside pressure and more dependent on their own internal structure. But this does not imply that they are any less strong. It must also be remembered that the quality of family life today is considered a matter of public concern, sufficiently so that the State and society at large feel entitled to look in and complain or take action if necessary. In a sense we are subjecting family life to a detailed factual scrutiny which would have been unthinkable, for example, to the Victorians. Obviously the result of such vigilance is to reveal the faults. But because we did not look before, we cannot assume that these faults were not there. However, it does seem that the modern family, as a social institution, stands reasonably clear and firm today, approved of by its members (as rising marriage rates, falling age of marriages and a rising birth rate attest) and supported by the State through a network of financial and social services.

Notes

1. O. R. McGregor, 'The Stability of the Family in the Welfare State', *Political Quarterly*, vol. 31, 1960.

2. R. M. MacIver and C. H. Page, *Society* (Macmillan, 1949).

3. T. Parsons, 'General Theory in Sociology', in R. K. Merton *et al.*, *Sociology Today* (New York, 1959).

4. B. Wilson, 'The Teachers' Role', *British J. of Sociology*, vol. 13 1962.

5. N. Anderson, *The Urban Community* (Routledge & Kegan Paul, 1960).

6. R. Fletcher, *The Family and Marriage* (Penguin Books, 1962).

7. C. C. Zimmerman, *The Family and Civilisation* (Harper, New York, 1947).

8. G. Rowntree and N. H. Carrier, 'The Resort to Divorce in England and Wales, 1858–1957', *Population Studies*, vol. 11, 1958.

9. S. Yudkin and A. Holme, *Working Mothers and Their Children* (Joseph, 1963).

10. V. Klein, *Working Wives* (Institute of Personnel Management, Occasional Papers, 15, 1958).

An activity is productive in the economic sense only if it makes use of scarce resources in such a way as to satisfy wants. A commodity capable of satisfying wants is said to have 'utility', and the terms on which it can be exchanged for other commodities depend both on its utility and on its scarcity. The connection between utility and market demand is discussed in the first extract from Alfred Marshall's 'Principles of Economics'. (Marshall, one of the greatest English economists, was Professor of Political Economy at the University of Cambridge from 1885 to 1908, and the 'Principles' were first published in 1890.)

Economists have traditionally accepted the state of the community's wants as a 'datum', and assumed that the performance of the economy was to be judged by the degree to which wants were satisfied – whether or not the wants sprang from tastes which the economists themselves shared. A different perspective is adopted by Professor J. K. Galbraith in the second extract, from his 'The Affluent Society'.

UTILITY AND DEMAND

Alfred Marshall

MAN cannot create material things. In the mental and moral world indeed he may produce new ideas; but when he is said to produce material things, he really only produces utilities; or in other words, his efforts and sacrifices result in changing the form or arrangement of matter to adapt it better for the satisfaction of wants. All that he can do in the physical world is either to readjust matter so as to make it more useful, as when he makes a log of wood into a table; or to put it in the way of being made more useful by nature, as when he puts seed where the forces of nature will make it burst out into life.

It is sometimes said that traders do not produce: that while the cabinet-maker produces furniture, the furniture-dealer merely sells what is already produced. But there is no scientific foundation for the distinction. They both produce utilities, and neither of them can do more: the furniture-dealer moves and rearranges matter so as to make it more serviceable than it was before, and the carpenter does nothing more. The sailor or the railway-man who carries coal above ground produces it, just as much as the miner who carries it underground; the dealer in fish helps to move on fish from where it is of comparatively little use to where it is of greater use, and the fisherman does no more. . . .

Utility is taken to be correlative to Desire or Want. . . . There is an endless variety of wants, but there is a limit to each separate want. This familiar and fundamental tendency of human nature may be stated in the *law of satiable wants* or *of diminishing utility* thus: The *total utility* of a thing to anyone (that is, the total pleasure or other benefit it yields him) increases with every increase in his stock of it, but not as fast as his stock increases. If his stock of it increases at a uniform rate the benefit derived from it increases at a diminishing rate. In other words, the additional benefit which a person derives from a given increase of his stock of a thing, diminishes with every increase in the stock that he already has.

That part of the thing which he is only just induced to purchase may be called his *marginal purchase*, because he is on the margin of doubt whether it is worth his while to incur the outlay required to obtain it. And the utility of his marginal purchase may be called the *marginal utility* of the thing to him. Or, if instead of buying it, he makes the thing himself, then its marginal utility is the utility of that part which he thinks it only just worth his while to make. And thus the law just given may be worded:

The marginal utility of a thing to anyone diminishes with every increase in the amount of it he already has.

There is, however, an implicit condition in this law which should be made clear. It is that we do not suppose time to be allowed for any alteration in the character or tastes of the man himself. It is therefore no exception to the law that the more good music a man hears, the stronger is his taste for it likely to become; that avarice and ambition are often insatiable; or that the virtue of cleanliness and the vice of drunkenness alike grow on what

they feed upon. For in such cases our observations range over some period of time; and the man is not the same at the beginning as at the end of it. If we take a man as he is, without allowing time for any change in his character, the marginal utility of a thing to him diminishes steadily with every increase in his supply of it.

Now let us translate this law of diminishing utility into terms of price. . . . Suppose, for instance, that tea of a certain quality is to be had at 2*s* per lb. A person might be willing to give 10*s* for a single pound once a year rather than go without it altogether; while if he could have any amount of it for nothing he would perhaps not care to use more than 30 lb in the year. But as it is, he buys perhaps 10 lb in the year; that is to say, the difference between the satisfaction which he gets from buying 9 lb and 10 lb is enough for him to be willing to pay 2*s* for it: while the fact that he does not buy an eleventh pound, shows that he does not think that it would be worth an extra 2*s* to him. . . . And our law may be worded:

The larger the amount of a thing that a person has, the less (other things being equal, i.e. the purchasing power of money, and the amount of money at his command, being 'given') will be the price which he will pay for a little more of it. . . .

A greater utility will be required to induce him to buy a thing if he is poor than if he is rich. . . . The richer a man becomes, the less is the marginal utility of money to him; every increase in his resources increases the price which he is willing to pay for any given benefit. And in the same way every diminution of his resources increases the marginal utility of money to him, and diminishes the price that he is willing to pay for any benefit. . . .

In large markets – where rich and poor, old and young, men and women, persons of all varieties of tastes, temperaments and occupations are mingled together – the peculiarities in the wants of individuals will compensate one another in a comparatively regular gradation of total demand. Every fall, however slight, in the price of a commodity in general use will, other things being equal, increase the total sales of it. . . .

There is, then, one general *law of demand*: The greater the amount to be sold, the smaller must be the price at which it is offered in order that it may find purchasers; or, in other words, the amount demanded increases with a fall in price, and diminishes with a rise in price.

THE DEPENDENCE EFFECT

J. K. Galbraith

I

THE notion that wants do not become less urgent the more amply the individual is supplied is broadly repugnant to common sense. It is something to be believed only by those who wish to believe. Yet the conventional wisdom must be tackled on its own terrain. Intertemporal comparisons of an individual's state of mind do rest on doubtful grounds. Who can say for sure that the deprivation which afflicts him with hunger is more painful than the deprivation which afflicts him with envy of his neighbour's new car? In the time that has passed since he was poor his soul may have become subject to a new and deeper searing. And where a society is concerned, comparisons between marginal satisfactions when it is poor and those when it is affluent will involve not only the same individual at different times but different individuals at different times. The scholar who wishes to believe that with increasing affluence there is no reduction in the urgency of desires and goods is not without points for debate. However plausible the case against him, it cannot be proved. In the defence of the conventional wisdom this amounts almost to invulnerability.

However, there is a flaw in the case. If the individual's wants are to be urgent they must be original with himself. They cannot be urgent if they must be contrived for him. And above all they must not be contrived by the process of production by which they are satisfied. For this means that the whole case for the urgency of production, based on the urgency of wants, falls to the ground. One cannot defend production as satisfying wants if that production creates the wants.

Were it so that a man on arising each morning was assailed by demons which instilled in him a passion sometimes for silk shirts, sometimes for kitchenware, sometimes for chamber-pots and sometimes for orange squash, there would be every reason to applaud the effort to find the goods, however odd, that quenched this flame. But should it be that his passion was the result of his first having cultivated the demons, and should it also be that his effort to allay it stirred the demons to even greater and greater

effort, there would be question as to how rational was his solution. Unless restrained by conventional attitudes, he might wonder if the solution lay with more goods or fewer demons.

So it is that if production creates the wants it seeks to satisfy, or if the wants emerge *pari passu* with the production, then the urgency of the wants can no longer be used to defend the urgency of the production. Production only fills a void that it has itself created.

II

The point is so central that it must be pressed. Consumer wants can have bizarre, frivolous or even immoral origins, and an admirable case can still be made for a society that seeks to satisfy them. But the case cannot stand if it is the process of satisfying wants that creates the wants. For then the individual who urges the importance of production to satisfy these wants is precisely in the position of the onlooker who applauds the efforts of the squirrel to keep abreast of the wheel that is propelled by his own efforts.

That wants are, in fact, the fruit of production will now be denied by few serious scholars. And a considerable number of economists, though not always in full knowledge of the implications, have conceded the point. . . . Keynes noted that needs of 'the second class', i.e. those that are the result of efforts to keep abreast or ahead of one's fellow being, 'may indeed be insatiable; for the higher the general level the higher still are they'.[1] And emulation has always played a considerable role in the views of other economists of want creation. One man's consumption becomes his neighbour's wish. This already means that the process by which wants are satisfied is also the process by which wants are created. The more wants that are satisfied the more new ones are born.

However, the argument has been carried farther. A leading modern theorist of consumer behaviour, Professor Duesenberry, has stated explicitly that 'ours is a society in which one of the principal social goals is a higher standard of living. . . . [This] has great significance for the theory of consumption . . . the desire to get superior goods takes on a life of its own. It provides a drive to higher expenditure which may even be stronger than that arising out of the needs which are supposed to be satisfied by that expenditure.'[2] The implications of this view are impressive. The notion of independently established need now sinks into the background. Because the society sets great store by ability to produce a high living standard, it evaluates people by the products they possess. The urge to consume is fathered by the value system which emphasises the ability of the society to produce. The more that is produced the more that must be owned in order to maintain the appropriate prestige. The latter is an important point, for, without going as far as Duesenberry in reducing goods to the role of symbols of prestige in the affluent society, it is plain that his argument fully implies that the production of goods creates the wants that the goods are presumed to satisfy.

III

The even more direct link between production and wants is provided by the institutions of modern advertising and salesmanship. These cannot be reconciled with the notion of independently determined desires, for their central function is to create desires – to bring into being wants that previously did not exist.[3] This is accomplished by the producer of the goods or at his behest. A broad empirical relationship exists between what is spent on production of consumers' goods and what is spent in synthesising the desires for that production. A new consumer product must be introduced with a suitable advertising campaign to arouse an interest in it. The path for an expansion of output must be paved by a suitable expansion in the advertising budget. Outlays for the manufacturing of a product are not more important in the strategy of modern business enterprise than outlays for the manufacturing of demand for the product. None of this is novel. All would be regarded as elementary by the most retarded student in the nation's most primitive school of business administration. The cost of this want formation is formidable. In 1956 total advertising expenditure – though, as noted, not all of it may be assigned to the synthesis of wants – amounted to about ten thousand million dollars. For some years it had been increasing at a rate in excess of a thousand million dollars a year. Obviously, such outlays must be integrated with the theory of consumer demand. They are too big to be ignored.

But such integration means recognising that wants are dependent on production. It accords to the producer the function both of making the goods and of making the desires for them. It recog-

nises that production, not only passively through emulation but actively through advertising and related activities, creates the wants it seeks to satisfy.

The businessman and the lay reader will be puzzled over the emphasis which I give to a seemingly obvious point. The point is indeed obvious. But it is one which, to a singular degree, economists have resisted. They have sensed, as the layman does not, the damage to established ideas which lurks in these relationships. As a result, incredibly, they have closed their eyes (and ears) to the most obtrusive of all economic phenomena, namely modern want creation.

This is not to say that the evidence affirming the dependence of wants on advertising has been entirely ignored. It is one reason why advertising has so long been regarded with such uneasiness by economists. Here is something which cannot be accommodated easily to existing theory. More pervious scholars have speculated on the urgency of desires which are so obviously the fruit of such expensively contrived campaigns for popular attention. Is a new breakfast cereal or detergent so much wanted if so much must be spent to compel in the consumer the sense of want? But there has been little tendency to go on to examine the implications of this for the theory of consumer demand and even less for the importance of production and productive efficiency. These have remained sacrosanct. More often the uneasiness has been manifested in a general disapproval of advertising and advertising men, leading to the occasional suggestion that they shouldn't exist. Such suggestions have usually been ill received.

And so the notion of independently determined wants still survives. In the face of all the forces of modern salesmanship it still rules, almost undefiled, in the textbooks. And it still remains the economist's mission – and on few matters is the pedagogy so firm – to seek unquestioningly the means for filling these wants. This being so, production remains of prime urgency. We have here, perhaps, the ultimate triumph of the conventional wisdom in its resistance to the evidence of the eyes. To equal it one must imagine a humanitarian who was long ago persuaded of the grievious shortage of hospital facilities in the town. He continues to importune the passers-by for money for more beds and refuses to notice that the town doctor is deftly knocking over pedestrians with his car to keep up the occupancy.

And in unravelling the complex we should always be careful not to overlook the obvious. The fact that wants can be synthesised by advertising, catalysed by salesmanship and shaped by the discreet manipulations of the persuaders shows that they are not very urgent. A man who is hungry need never be told of his need for food. If he is inspired by his appetite, he is immune to the influence of Messrs Batten, Barton, Durstine and Osborn. The latter are effective only with those who are so far removed from physical want that they do not already know what they want. In this state alone men are open to persuasion.

IV

The general conclusion of these pages is of such importance for this essay that it had perhaps best be put with some formality. As a society becomes increasingly affluent, wants are increasingly created by the process by which they are satisfied. This may operate passively. Increases in consumption, the counterpart of increases in production, act by suggestion or emulation to create wants. Or producers may proceed actively to create wants through advertising and salesmanship. Wants thus come to depend on output. In technical terms it can no longer be assumed that welfare is greater at an all-round higher level of production than at a lower one. It may be the same. The higher level of production has, merely, a higher level of want creation necessitating a higher level of want satisfaction. There will be frequent occasion to refer to the way wants depend on the process by which they are satisfied. It will be convenient to call it the Dependence Effect.

We may now contemplate briefly the conclusions to which this analysis has brought us.

Plainly the theory of consumer demand is a peculiarly treacherous friend of the present goals of economics. At first glance it seems to defend the continuing urgency of production and our preoccupation with it as a goal. The economist does not enter into the dubious moral arguments about the importance or virtue of the wants to be satisfied. He doesn't pretend to compare mental states of the same or different people at different times and to suggest that one is less urgent than another. The desire is there. That for him is sufficient. He sets about in a workmanlike way to satisfy

desire, and accordingly he sets the proper store by the production that does. Like woman's his work is never done.

But this rationalisation, handsomely though it seems to serve, turns destructively on those who advance it once it is conceded that wants are themselves both passively and deliberately the fruits of the process by which they are satisfied. Then the production of goods satisfies the wants that the consumption of these goods creates or that the producers of goods synthesise. Production induces more wants and the need for more production. So far, in a major *tour de force*, the implications have been ignored. But this obviously is a perilous solution. It cannot long survive discussion.

Among the many models of the good society no one has urged the squirrel wheel. Moreover, as we shall see presently, the wheel is not one that revolves with perfect smoothness. Aside from its dubious cultural charm, there are serious structural weaknesses which may one day embarrass us. For the moment, however, it is sufficient to reflect on the difficult terrain which we are traversing. We have seen how deeply we were committed to production for reasons of economic security. Not the goods but the employment provided by their production was the thing by which we set ultimate store. Now we find our concern for goods further undermined. It does not arise in spontaneous consumer need. Rather, the dependence effect means that it grows out of the process of production itself. If production is to increase, the wants must be effectively contrived. In the absence of the contrivance the increase would not occur. That is not true of all goods, but that it is true of a substantial part is sufficient. It means that since the demand for this part would not exist, were it not contrived, its utility or urgency, ex contrivance, is zero. If we regard this production as marginal, we may say that the marginal utility of present aggregate output, ex advertising and salesmanship, is zero. Clearly the attitudes and values which make production the central achievement of our society have some exceptionally twisted roots.

Perhaps the thing most evident of all is how new and varied become the problems we must ponder when we break the nexus with the work of Ricardo and face the economics of affluence of the world in which we live. It is easy to see why the conventional wisdom resists so stoutly such change. It is a far, far better thing to have a firm anchor in nonsense than to put out on the troubled seas of thought.

Notes

1. J. M. Keynes, 'Economic Possibilities for our Grandchildren', in *Essays in Persuasion* (London, 1933), pp. 365–6.

2. J. S. Duesenberry, *Income, Saving and the Theory of Consumer Behaviour* (Cambridge, Mass., 1949).

3. Advertising is not a simple phenomenon. It is also important in competitive strategy and want creation is, ordinarily, a complementary result of efforts to shift the demand curve of the individual firm at the expense of others or (less importantly, I think) to change its shape by increasing the degree of product differentiation. Some of the failure of economists to identify advertising with want creation may be attributed to the undue attention that its use in purely competitive strategy has attracted. It should be noted, however, that the competitive manipulation of consumer desire is only possible, at least on any appreciable scale, when such need is not strongly felt.

Given the objectives of economic activity, the problem is to allocate scarce resources in such a way as to attain these objectives as efficiently as possible. In the first essay which follows, Ralph Turvey discusses how such allocations can be chosen. It has traditionally been assumed that competition will force producers to adopt efficient methods of production and keep prices at the minimum level needed to call forth the output needed to satisfy demand; monopolies, and other restraints on competition, bring about less efficient allocations of resources. A powerful argument against this view is put forward in the second extract by J. A. Schumpeter, one of the most distinguished of twentieth-century economists.

HOW RATIONAL ECONOMIC DECISIONS ARE MADE

Ralph Turvey

THIS article is about the most rational way of making economic decisions where the problems under review relate to a part of the economy, rather than to the whole. Whether London should have a third airport and, if so, whether Stansted is the right choice is one example; the best way of utilising natural gas from the North Sea is another. Such problems involve particular uses of particular resources, as opposed to the general functioning of the economy where economic aggregates like unemployment and the balance of trade are involved.

The kind of policy problem discussed here involves a choice between a number of specific alternatives. What has to be done, first, is to sort out the possibilities open and, then, to decide which of them is the best. Thus, the search for a solution to London's traffic problem requires setting down and then comparing the various alternatives. The first part of this, listing and describing the things that could be done, will not be discussed here. I am concerned in this article with the second stage, that of choice. The argument is that better choices will be made if the comparison is carried out in a formalised, quantitative manner than if it is not.

CHOICE AND MAXIMISATION

The argument starts with the proposition that consistent choice between alternatives can be described as maximisation subject to constraints. This proposition is, of course, really true by definition. A simple example will illustrate its meaning. Take the choice of a holiday. It can be represented as the maximisation of pleasure, subject to a time constraint (say three weeks), a foreign-currency constraint (£50 per head) and so on. Now, this is obviously not really a useful way of deciding how to plan a holiday. The example is given merely in order to show that the proposition is a completely general one. Once this is accepted, the issue becomes one of where it is useful. The purpose of this article is to show that it can be employed in dealing with a wide range of economic policy problems.

The formulation of these problems in terms of maximising something, subject to a number of constraints, serves two purposes. First, it aids communication by presenting the relevant factors clearly and systematically, so that people involved in making a decision can share a common frame of reference. At the worst, where they do not share such a common frame, at least it enables them to isolate the sources of disagreement. Second, in a restricted range of cases, it may provide a means of formulating unequivocal rules for making decisions.

What is involved, then, is the application of certain modern scientific management techniques to issues of economic policy, so as to illuminate and clarify them and – occasionally – to provide a direct answer. These techniques, as applied to economic policy, go under such names as 'cost-benefit analysis', 'cost-effectiveness budgeting', 'systems analysis' or '*le calcul économique*'. Their essence is an attempt to quantify the arguments for and against the alternatives considered.

The thing which is to be maximised is called the 'objective function' or 'social pay-off', and it has to be expressed in terms of some common denominator in order that its components can be added. The best such common denominator we have is money; indeed, it is probably the only one which is usable in practical

problems. Thus, in tackling economic policy problems, the starting point is to make a list of gains and losses expressed in terms of money. The objective function is the sum of these amounts, with gains as positive items and losses (costs) as negative items. Occasionally, the problem is limited to choosing the best way of achieving some given result and, in such cases, the objective function can be defined as the monetary value of the social costs which then has to be minimised. This is just a matter of convenience in exposition, however, for minimising losses minus gains is the same as maximising gains minus losses.

An important requirement for the list of gains and losses entering into the objective function is that they shall be mutually exclusive, that is, there is no double counting. Thus, the cost of agricultural land taken for a development scheme and the loss of its agricultural production are two facets of the same loss and they cannot both be included. Another requirement is that the list shall be as exhaustive as possible, but gains or losses which are common to all the alternatives being considered can be left out, since it is only the differences between them which are relevant to choice. Thus, *if* Stansted and Sheppey were equally suitable as an airport from the point of view of the airlines, and *if* one or the other had to be chosen, airline gains and losses could be left out of the calculations.

NATURE OF CONSTRAINTS

Before discussing any further the listing and evaluation of items in the objective function, let us consider briefly the constraints relevant to maximising it in economic policy problems of the kind mentioned. Apart from those whose significance is primarily computational, they fall into two main classes.

The first class consists of limits that exist on the policy-maker's freedom of choice. For example, if total capital expenditure cannot exceed a certain limit, then that limit is such a constraint. Physical limitations may play a similar role: there may be some maximum annual rate at which houses can be built in one neighbourhood of a new town or there may be an unavoidable bottleneck in the supply of some key equipment.

The other class of constraints brings us to those gains and losses which cannot be included in the analysis because they cannot be expressed in monetary terms. A requirement that a particular landscape shall not be spoiled provides an extreme example of this. Another is a condition that, for social reasons, employment in a particular area shall not fall below a certain figure. Such constraints thus express the maximum loss tolerable (or minimum gain acceptable) as limits. Because such limits relate to gains and losses which are incapable of measurement and valuation, they can hardly fail to be arbitrary. This arbitrariness, however, is inherent in the problem being tackled and is in no way generated by the method of analysis. It is, indeed, a positive merit of this method that it brings out so explicitly the need for judgements to be made concerning relevant factors which cannot be quantified. That something cannot be measured in monetary terms does not make it any less important than something which can.

The approach to decision-making which this paper advocates has now been set out in general terms. In fact, this is really no more than a systematic way of setting out the pros and cons of the alternative proposals which have to be considered. It thus forces the proponents of different schemes to deploy their arguments within a rigorous logical framework; if it does not produce agreement it at least identifies the sources of disagreement. Furthermore, it sometimes has the incidental advantage of suggesting schemes that might otherwise have been overlooked. But the main advantage is more fundamental: namely, the quantification of as many of the arguments as possible, so that the relative weight to be given to each is explicitly assessed. Measurement replaces mere advocacy.

The extent to which the solution to any particular problem demands guess and judgement, as well as calculation, varies from case to case. It would be naïve to pretend that science takes over from judgement. Nevertheless, it is equally naïve to reject the measurement of some factors because other factors cannot be measured, or not to prefer a rough measurement to the absence of any measurement. The middle path is to welcome an improved approach to decision-making without expecting perfection, to see what can be done along these lines in each particular case, to prefer a little knowledge to mere hunch.

TWO EXAMPLES

The general description of this method of making rational decisions given so far is rather vague. An adequate treatment of any

real case, however, would need far more space than is available here, so two hypothetical examples will have to serve. Their purpose is merely to show the broad way in which the method can be applied, and they do *not* purport to do anything more than this.

Suppose, first, that in a certain town the growth of water consumption threatens to overtake the capacity of the water supply system. Suppose, too, that three alternatives exist:

1. Capital expenditure on a scheme to increase water supply capacity.
2. Closing down a factory which is a particularly large water user.
3. Charging such a price per gallon for water (which involves the introduction of meters) as is necessary to restrain the growth of demand within the limits of the existing capacity.

The objective function appropriate for comparing these three will include:

The value to consumers of any additional water consumption;
minus any capital expenditure to increase capacity;
plus the value in alternative uses of the factory space, labour, etc. (net of reconversion and redeployment costs) if the factory closes down;
minus the value added to the output of the factory if it closes down;
minus any cost incurred in installing and operating water meters.

Only some of these items will come into the calculations for each of the three alternatives, as follows:

1. Value to all consumers of additional water consumption;
minus capital expenditure to increase capacity.
2. Value in alternative uses of the factory space, labour, etc. (net of reconversion and redeployment costs);
minus value added of the output of the factory;
plus value to all other consumers of their additional water consumption.
3. *minus* the cost of installing and operating meters.

Note that the absence of any increase in water consumption is not included as a cost in alternative (3), since the inclusion of such increases in cases (1) and (2) brings it into the comparison.

This simplified example brings out two important points. First, the various items have a time dimension. Thus, if the supply system were extended in order to provide more water, the capital expenditure would cover a year or two and then be followed, year after year, by the benefit to consumers of the additional supplies. Reduction to a common denominator of the various items in the objective function therefore requires not only the use of money as a measure, but also the discounting of future benefits and costs to make them comparable with immediate gains and losses. We must introduce into the calculation a discount rate which expresses the responsible decision-taker's choice between bread today and jam tomorrow.

The second point relates to an apparent omission. Suppose that the objective function is maximised by introducing meters and charging for water on a per gallon basis. All water users will now lose through having to pay more for their water consumption, which is now rationed by price to prevent any increase. Why does the objective function not include their loss? The answer is that the matter is viewed from the national viewpoint; the consumers' loss is the water undertaking's gain, so together they cancel out. Such 'transfers' need be reckoned only if they are from the 'deserving' to the 'undeserving', or vice versa. Since any advantage or disadvantage derived from them cannot be measured, being of a social or political nature, the analyst cannot put it in the objective function. It must, nevertheless, be brought to the decision-maker's attention if it is his legitimate concern, possibly being incorporated in a constraint.

As explained above, this example is not complete. But it does show what sort of item has to go into an objective function and how alternatives can be compared. In this case, the value of the function might be negative in all three alternatives, so that maximising it means minimising a negative sum – choosing, that is, the alternative with the lowest cost.

As a second example, we may consider a project for smelting aluminium in Britain using a nuclear reactor built to provide the necessary power, so that more alumina and less aluminium are imported. Such a project would involve several parties, but its

examination from the point of view of the country as a whole would also seem necessary. Only if gains outweigh losses for the nation can there be an excess of gains over losses for all of the parties concerned. A comprehensive calculation in terms of the social interest is thus required. Looked at in this way, the two alternatives are: Do nothing; or, Go ahead. The objective function for the first is, of course, zero. The main items in that for the second alternative are:

The cost saving from importing alumina instead of the equivalent quantity of aluminium;
minus construction cost of the nuclear plant (including initial fuel charge) and the smelter;
minus running and maintenance costs of the nuclear plant (including refuelling) and the smelter;
minus cost of provision of standby supply by the electricity industry.

These all have to be expressed in present-worth terms. (Note that the running cost of the smelter excludes electricity costs, since these are already included.) If the smelter and nuclear plant were separately owned, the price paid for the power provided would affect the division of the gain (or loss) between the two owners, but this is irrelevant to the magnitude of the total gain (or loss) seen from the national viewpoint.

PRIVATE AND PUBLIC INTERESTS

If a private business is run with the sole aim of maximising profits, its objective function is simply revenue less costs. The objective functions appropriate for economic policy decisions taken on grounds of national interest may clearly differ from this; and the constraints which are relevant to any particular problem viewed in terms of the public interest may also differ from those applicable to a purely business decision.

There are four groups of circumstances that can make business revenue and costs a poor measure of gains and losses from a national viewpoint. The most obvious is that the output or services provided may not be marketed. Thus, a road improvement yields no direct revenue proportionate to its usefulness; neither do coastal protection works. In such cases, the value of the project to its beneficiaries has to be estimated.

Second, some of the gains or losses may accrue to third parties. A classic case of this is the benefit road users in central London are expected to get from the new underground Victoria Line: the diversion of traffic to it will ease congestion elsewhere. Another example is to be found in salmon fisheries, where a change in the type or extent of fishing at river mouths affects rod fishing upstream.

Third, the prices received for output or paid for 'inputs' may fail to measure national gains and costs because of the way some part of the economy functions. The market value of agricultural land, for instance, is affected by agricultural subsidies. The price of motor fuel contains a large tax element, i.e. a transfer. The cost to the economy of employing in a development area someone who would otherwise be unemployed is zero, but the employer incurs the cost of wages and insurance contributions. In such cases, it may be necessary to impute a 'shadow price', that is, to correct the ruling market price to obtain a social valuation.

Another example of the need to adjust market prices in estimating gains and losses might be thought to arise in the case of the project for an aluminium smelter. If it were urged in favour of the project that it would make a substantial contribution to the balance of payments, one would have to ask whether the first item in the calculation outlined above does not already take this into account. Anyone who denies this is, in effect, saying that each \$2.40 saving of foreign exchange is worth more than £1 of domestic resources. This could be reflected in the calculation by valuing the import saving at more than £1 per \$2.40.[1] Thus, the great merit of the approach is shown once again: it forces the use of an explicit quantitative evaluation, instead of a loose appeal to 'relevant considerations'.

The final circumstance which can make business revenue and costs poor measures of national gain and loss arises when a proposed scheme makes such a large impact upon a market that the ruling market price is affected by it. We then have to decide whether the old or the new price, or some combination, provides the relevant valuation for insertion in the objective function.

In all these four groups of circumstances, the problem of estimating value in some indirect manner has to be tackled. Value

has no mystical meaning here: it is simply the monetary expression of the sacrifice which people are prepared to make in order to obtain something. Conceptually, this can be explained in terms of the logical proposition with which we started. Consistent choice, it was pointed out, can be described as the maximisation of an objective function subject to constraints. It follows that, where choice can be observed and where it is consistent, the value placed by the decision-maker upon some of the items in the objective function can be inferred. Thus, the fundamental logical point is, in a sense, reversible. Given sufficient data, either choice can be deduced from values, or values deduced from choice.

The problem, then, is not conceptual but practical and the feasibility of estimating value in any particular case depends entirely upon its individual circumstances. No estimate will ever be completely reliable but a poor estimate is better than none at all, unless the cost of the necessary research is large.

Consider, to go back to our example, the value to consumers of the additional water supply to be provided by an extension of supply capacity. Suppose that no extra water is provided and that a per gallon charge is introduced to cut down the use of water to existing supply capacity. If this hypothetical charge has to be x pence per thousand gallons, the free choice of all users at this price implies that this is the value to all of them of a little more or little less water. Thus, if they value it at more they would wish to use more and the price which rations out the existing supply would have to exceed x pence per thousand gallons.

Suppose, alternatively, that the extra water is provided, with water rates continuing to be independent of water use, so that the cost to each user of extra consumption is zero. The same argument shows that in this case the value to all of them of a little more or less water is also zero.

On average, therefore, the extra water is worth something between x pence and zero per thousand gallons: $\frac{1}{2}x$ per thousand gallons is thus a first estimate (capable of refinement) of the value of the extra water. This means that the estimate requires knowledge of the size of x, that is, of the response of water users to the imposition of a charge related to the amount of water used. Such knowledge is attainable, though at a cost. The necessary research would consist of comparing the consumption of a group of water users who pay for metered water supplies with that of a 'control group' who do not, and whose use of water would be measured.

Once again, the example is simplified. But, even so, it shows how modern techniques can be used so that decisions are made on a more rational basis. Irreducible uncertainty, of course, remains, but the approach outlined above can, nevertheless, improve upon rough and ready hunches.

Note

1. The cost of imported nuclear fuel would also have to be corrected.

THE PROCESS OF CREATIVE DESTRUCTION

J. A. Schumpeter

THE fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organisation that capitalist enterprise creates.

The contents of the labourer's budget, say from 1760 to 1940, did not simply grow on unchanging lines but they underwent a process of qualitative change. Similarly, the history of the productive apparatus of a typical farm, from the beginnings of the rationalisation of crop rotation, ploughing and fattening to the mechanised thing of today – linking up with elevators and railroads – is a history of revolutions. So is the history of the productive apparatus of the iron and steel industry from the charcoal furnace to our own type of furnace, or the history of the apparatus of power production from the overshot water wheel to the modern power plant, or the history of transportation from the mailcoach to the aeroplane. The opening-up of new markets, foreign or domestic, and the organisational development from the craft shop and factory to such concerns as U.S. Steel illustrate the same process of industrial mutation – if I may use that biological term – that incessantly revolutionises[1] the economic structure *from within*, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism. It is what capitalism consists in and what every capitalist concern has got to live in.

Since we are dealing with an organic process, analysis of what happens in any particular part of it – say, in an individual concern or industry – may indeed clarify details of mechanism but is inconclusive beyond that. Every piece of business strategy acquires its true significance only against the background of that process and within the situation created by it. It must be seen in its role in the perennial gale of creative destruction; it cannot be understood irrespective of it or, in fact, on the hypothesis that there is a perennial lull.

But economists who, *ex visu* of a point of time, look for example at the behaviour of an oligopolist industry – an industry which consists of a few big firms – and observe the well-known moves and countermoves within it that seem to aim at nothing but high prices and restrictions of output are making precisely that hypothesis. They accept the data of the momentary situation as if there were no past or future to it and think that they have understood what there is to understand if they interpret the behaviour of those firms by means of the principle of maximising profits with reference to those data. The usual theorist's paper and the usual government commission's report practically never try to see that behaviour, on the one hand, as a result of a piece of past history and, on the other hand, as an attempt to deal with a situation that is sure to change presently – as an attempt by those firms to keep on their feet, on ground that is slipping away from under them. In other words, the problem that is usually being visualised is how capitalism administers existing structures, whereas the relevant problem is how it creates and destroys them. As long as this is not recognised, the investigator does a meaningless job. As soon as it is recognised, his outlook on capitalist practice and its social results changes considerably.[2]

The first thing to go is the traditional conception of the *modus operandi* of competition. Economists are at long last emerging from the stage in which price competition was all they saw. As soon as quality competition and sales effort are admitted into the sacred precincts of theory, the price variable is ousted from its dominant position. However, it is still competition within a rigid pattern of invariant conditions, methods of production and forms, of industrial organisation in particular, that practically monopolises attention. But in capitalist reality as distinguished from its textbook picture, it is not that kind of competition which counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organisation (the largest-scale unit of control for instance) – competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives. This kind of competition is as much more effective than the other as a bombardment is in comparison with forcing a door, and so much more important that it becomes a matter of comparative indifference whether competition in the ordinary sense functions more or less promptly; the powerful lever that in the long run expands output and brings down prices is in any case made of other stuff.

It is hardly necessary to point out that competition of the kind we now have in mind acts not only when in being but also when it is merely an ever-present threat. It disciplines before it attacks. The businessman feels himself to be in a competitive situation even if he is alone in his field or if, though not alone, he holds a position such that investigating government experts fail to see any effective competition between him and any other firms in the same or a neighbouring field and in consequence conclude that his talk, under examination, about his competitive sorrows is all make-believe. In many cases, though not in all, this will in the long run enforce behaviour very similar to the perfectly competitive pattern.

Many theorists take the opposite view which is best conveyed by an example. Let us assume that there is a certain number of retailers in a neighbourhood who try to improve their relative position by service and 'atmosphere' but avoid price competition and stick as to methods to the local tradition – a picture of stagnating routine. As others drift into the trade that quasi-equilibrium is indeed upset, but in a manner that does not benefit their customers. The economic space around each of the shops having been narrowed, their owners will no longer be able to make a living and they will try to mend the case by raising prices in tacit agreement. This will further reduce their sales and so, by successive pyramiding, a situation will evolve in which increasing potential supply will be attended by increasing instead of decreasing prices and by decreasing instead of increasing sales.

Such cases do occur, and it is right and proper to work them out. But as the practical instances usually given show, they are fringe-

end cases to be found mainly in the sectors farthest removed from all that is most characteristic of capitalist activity.[3] Moreover, they are transient by nature. In the case of retail trade the competition that matters arises not from additional shops of the same type, but from the department store, the chain store, the mail-order house and the supermarket which are bound to destroy those pyramids sooner or later.[4] Now a theoretical construction which neglects this essential element of the case neglects all that is most typically capitalist about it; even if correct in logic as well as in fact, it is like *Hamlet* without the Danish prince.

Notes

1. Those revolutions are not strictly incessant; they occur in discrete rushes which are separated from each other by spans of comparative quiet. The process as a whole works incessantly, however, in the sense that there always is either revolution or absorption of the results of revolution, both together forming what are known as business cycles.

2. It should be understood that it is only our appraisal of economic performance and not our moral judgement that can be so changed. Owing to its autonomy, moral approval or disapproval is entirely independent of our appraisal of social (or any other) results, unless we happen to adopt a moral system such as utilitarianism which makes moral approval and disapproval turn on them *ex definitione*.

3. This is also shown by a theorem we frequently meet with in expositions of the theory of imperfect competition, viz. the theorem that, under conditions of imperfect competition, producing or trading businesses tend to be irrationally small. Since imperfect competition is at the same time held to be an outstanding characteristic of modern industry, we are set to wondering what world these theorists live in, unless, as stated above, fringe-end cases are all they have in mind.

4. The mere threat of their attack cannot, in the particular conditions, environmental and personal, of small-scale retail trade, have its usual disciplining influence, for the small man is too much hampered by his cost structure and, however well he may manage within his inescapable limitations, he can never adapt himself to the methods of competitors who can afford to sell at the price at which he buys.

In a 'market economy', or 'price system', resources are directed towards the production of different commodities in accordance with the strength of consumers' demand for them. In the first extract below, Professors Phelps Brown and Wiseman describe the principles on which a price system operates.

An alternative method of allocating resources is through a 'command economy', in which a central authority has power to organise production in accordance with a comprehensive plan; the obvious example of such a system is, of course, the Soviet Union. In the second extract, Professor Nove describes the Soviet planning system and indicates some of the problems to which it gives rise.

THE PRICING SYSTEM AND RESOURCE ALLOCATION

E. H. Phelps Brown and J. Wiseman

THE WORKING OF A PRICING SYSTEM

THE functions which a pricing system can perform will be shown in a clear light if we begin with an economy in which there is no element of such a system, and then introduce one by degrees.

An economy in which there was no element of a pricing system would be governed either by custom or by central direction. In a completely customary economy, everyone would do the work, and receive the share of the produce, which custom prescribed, and his own preferences would have no influence either on what he did or on what he got. In a completely directed economy, the use to be made of each man's labour and of each piece of equipment, and the rations of produce to be issued to each man, would be laid down by a central authority. This authority might be appointed democratically, and the plan it made might be subject to discussion and amendment by the people; but it would have to be cast in general terms, and each man would have to do what he was told, and take what he was given, without having any means of bringing his own tastes to bear directly on his choice of job or of produce. An army provides an example of such an economy. Here, though an endeavour may be made to put each man into the kind of post for which he is best suited or which he would most like, the way in which men are distributed between different posts is not left to them to work out through each man's endeavour to get the kind of post that will suit him or pay him best: orders are orders, and men must do the jobs to which they are appointed. Nor, though ration scales will be worked out to meet the needs of particular jobs, or the tastes of the majority, will they be tailor-made to the needs or tastes of any one man; but each man will get the rations prescribed for the likes of him, and he cannot elect to take rather more of one thing on condition of giving up somewhat of another.

But we have seen [in Chapter 3] how prisoners of war develop a system of exchanges to adjust a standard issue to various personal needs: the non-smoker swaps his tobacco for the smoker's chocolate. A first modification of the completely directed economy towards meeting personal tastes would be effected if this system of redistribution were recognised and extended as it would be if each man were issued not with rations in kind but with so many points, which he could lay out as he chose on goods ticketed with the number of points he must give up for them. The non-smoker, then, instead of drawing tobacco which he later bartered for chocolate, would draw more chocolate in the first place, leaving the tobacco to be drawn by the smoker who for his part had left him the extra chocolate. There would be no assurance, however, that the sum of all intended takings would be just what was available for issue. At the requirements in points first ticketed, some goods might hang fire and accumulate in stock, while others were in short supply. This maladjustment could not be remedied, on our present assumptions, by having less of the first kind of good made, and more of the second, for we are still supposing that each unit of productive factor is assigned its role by a fixed plan. What can be done, however, is to alter the tickets, to reduce the points required for the goods that are hanging fire, and raise the points required

for the goods in short supply, until consumers laying out the same total of points as before have been induced to raise their takings of the first kind of good, and reduce those of the second, to equality with the available supplies. In this way the same products as before would be distributed among the consumers in a way which was influenced by the personal tastes of each. The number of points to be issued to each man would be decided as part of the plan. Given this, each man would be free to take what he wanted, subject to two conditions: that every other man should be free in the same way, and that the sums of their several demands should just equal the supplies available. The terms on which any good could be obtained would be responsive to the tastes of each and every consumer.

The obvious practical form of this mechanism is the payment of incomes in money, and the putting of prices on the goods. We have spoken here not of money but of points, because money has more functions than are required here. Money is not only a medium of exchange but a store of value: you can not only spend money on one thing or another, but you can choose not to spend it at all just now, and keep it to spend later on. The points, however, could have a period of currency after which they ceased to be valid. They would then have only the function of a medium of exchange, and this is all that is necessary for the present purpose. For simplicity we shall speak henceforward of money and prices, but it will be remembered that only one of the actual functions of money is relevant here.

A second modification of the directed economy towards meeting personal tastes would now be made, if it were noticed that a glut of one good and shortage of another might be corrected not only by changing their prices, but also by taking some productive factors away from making the first, and putting them on to make more of the second. This could be done in a rough and ready way if the appearance of any glut or shortage were made the signal for some reduction or increase of output. The first modification allowed the personal tastes of consumers to influence the distribution among them of given quantities of products. This second modification would carry such influence farther back, and allow it to affect the distribution of given productive resources between the making of different products. What is being organised now is a kind of consumers' referendum on how productive resources shall be allocated: each penny that a consumer spends will draw productive factors towards meeting his particular wants.

But the working of the second modification is more complicated than that of the first. If the appearance of a shortage calls for an increase in the output of some commodity, how big should the increase be? How much of the shortage should be met by raising output, and how much by raising price and checking demand? From which other lines of production should the additional productive factors be withdrawn? These problems can be solved by the same sort of device as in the market for consumers' goods. Corresponding to these goods now are the various productive factors available for hire, and corresponding to the consumers are the undertakings – let us call them firms, whether they are public or private – which combine and direct the services of the productive factors, and turn out the goods which are sold to consumers. Let prices be put on the productive factors, and let firms spend on them the revenue that they collect from the sale of their products. They will then be acting as the deputies of the consumers who buy their goods, in bidding for the services of productive factors to make those goods. With this mechanism there can be worked out the distribution of productive factors between different lines of output, which will give consumers the greatest satisfaction on the whole. The way in which this can be done requires separate consideration.

THE PRICING SYSTEM AND THE 'WELFARE OPTIMUM'

The argument of the last section provides us with a means of establishing certain characteristics of the system of competitive markets that are relevant to economic policy. It is these characteristics, we said in the last chapter, that provide the main approach of welfare economics. We can now survey them.

Let us assume for the present that we have no quarrel with the distribution of income between consumers, so that if they are prepared to pay more for X than Y, we shall infer that they will derive greater satisfaction on the whole if it is X rather than Y that is made available. Now with any given distribution of productive factors between different lines of output, we can always propose to switch some parcel of factors from one line to another, so that the output of one line is raised by X at the cost of lowering

the output of another line by Y. If consumers are prepared to pay more for X than Y, such a switch will be a change for the better. The best possible distribution will have been reached when no further switch of this kind remains possible.

Economic analysis shows us that in this state of affairs three conditions will be satisfied:

(*a*) The marginal value[1] product of each factor must be the same in all employments: otherwise a change for the better could be made by switching a unit of a factor from a use in which its marginal value product is lower to one in which it is higher.

(*b*) In each firm, the intakes of factors must be so adjusted that their marginal value products are proportional to their prices: otherwise the firm could get the same output for a lower total cost by hiring more of a factor whose marginal value product bore a relatively high ratio to its price, and less of some other for which the corresponding ratio was lower. The firm will have got the best distribution of its outlay on productive factors when the marginal pound spent on each factor brings the same addition to total value product. . . .

(*c*) Each firm must lay out on the hire of factors the same amount as it receives from the sale of their product, otherwise it will not faithfully transmit the intent of the consumers whose deputy it is. Generally, the greater the quantity of a product that is offered for sale at a given time, the lower is the price that will clear it. If we compare the total revenue from the sale of greater quantities with that from the sale of less, we find that, after a point, an increase in quantity is offset and more than offset by the fall in price it brings with it, and total revenue falls. But, again at any one time, the total cost of a bigger quantity is always greater than that of a smaller. There will be some quantity of output, then, which is too big, in the sense that the total revenue from its sale does not cover the total cost of making it; and some other quantity which is too small, in the sense that revenue exceeds cost.[2] Between them lies a quantity for which revenue equals cost, and what the consumers are willing to pay for the product is exactly passed on, neither more nor less, as their payment for the factors required to make it. The present condition says that this is the quantity that must be produced. Because it prescribes the scale of output, we may call it the scale rule. Its application varies, according to the duration and extent of the adjustments to which we apply it. In the short run, a firm can vary its output only by varying the intake of direct labour and materials in a plant and organisation of given size; in the longer run, it can change the size of the plant and organisation themselves. The output of a given product can be varied, not only by varying the outputs of the firms that make it, but also by varying the number of firms in the trade, or the division of labour between such firms: that is, the scale of the trade itself may be varied.

For our present purpose, however, we do not need to pursue these applications, but need only get hold of the scale rule itself. The function of this rule is to ensure, in the referendum by which consumers through their spending are voting on the distribution of productive factors between different uses, that the voting is not 'cooked' in the ballot boxes kept by the firms. If abnormal profits persist, then more is being spent by consumers on the output than is needed to buy in the factors required to make that output: the consumers are voting for a higher output, and the firm is stopping their vote from taking effect. If losses, on the other hand, are being sustained, then more is being paid out to retain factors in a given use than consumers are willing to pay for their product there. The scale rule precludes both kinds of distortion.

THE PRICING SYSTEM AS A FORM OF CONTINUOUS REFERENDUM

We have been looking at the pricing system as a device for reaching the best distribution of the productive resources of an economy between different possible uses. It works through two connected markets, the consumers' market and the market for productive factors. The first shows what valuations the consumers in the aggregate set upon various products. The second works back from these valuations to set derived valuations upon the factors that make the products: the price of a unit of a factor is equal to the consumers' valuation of the most valued addition which that unit can make, incrementally, to output.

We have already likened this process to a system of indirect

election: in bidding for the use of factors, firms are casting votes to decide the distribution of the resources of the economy, and these votes are proxies which consumers have lodged with them in buying their products. The daily outlay of the consumers thus provides, in the pure type of pricing system, a continuous referendum on how the resources of the economy shall be used. Who buys meat is voting for land, labour and equipment to be used for producing meat, and not be let go to produce something else. Who tries to buy more meat is voting for factors to be withdrawn from other uses and brought in to increase the output of meat. The firms which respond, according to the usual business reckoning of cost and revenue, to changes in consumers' buying from them, are passing those votes back, to decide the distribution of productive factors. The price which the market sets on a productive factor registers its power to give consumers what they want, and so its attractiveness to the voter. The rule that a firm must pay its way applies the principle that making a product is justifiable only if the factors required to make it cannot be used in other ways which consumers will value more: for the firm's receipts express the valuation that consumers put on its present products, and its costs express the valuation that they put on alternative uses of the factors it employs. A unit of a factor may be retained in one use only if the votes cast to keep it there are not fewer than those it could get elsewhere.

THE PRICING SYSTEM AS A DEVICE FOR ALLOCATING RESOURCES

The analogy of the referendum does not bring out two special features of the pricing system: it registers the strengths of different demands in a common scale; and while leaving men free to make their own choices, it induces them to adjust their demands themselves so that they shall be compatible with one another in the aggregate. The meaning of these features can be seen by contrast with the way in which a scarce resource is distributed between different claimants when the pricing system is not used.

Suppose that there is a limited amount of constructional steel available, and that instead of its being left to go to those who will pay most for it, it is to be allocated among different claimants by a government department. Demands on this steel will be made by those who want it for factories of various kinds, those who want it for private dwellings, and those who would use it in public buildings such as schools and health centres. In considering these demands, the department will encounter two difficulties: each claimant will plead his case, but there is no scale on which each will register the strength of his demand so that it can be compared with others; the total demanded will only by rare coincidence be equal to the total available, so that either the department will have to cut down the present demands, or it will have to try to call forth some new ones. The first difficulty, of course, would not bother the department were it not for the second: but as soon as the department has to adjust the total demanded to the total available, it needs to decide which of the present demands could bear some reduction most easily, or what extensions of demand would satisfy most want. If the claimants had been invited to tender for supplies, that is, to submit offers to buy a stated amount at a stated price, they could be arranged in the order of the prices they offered, and the available supply could be allotted as far down the list of claimants as it would go, beginning with the highest bidder; but in fact they are not asked to bid, but only to plead, and there is no common scale in which the strength of their demands is expressed and arranged in order. The department, then, must ask for a directive, a statement of the weight which the government attaches to different types of claim, and this directive is likely to take the form of a statement of priorities – that the builders of factories, for instance, shall be given first claim, and that, in allocating what is left, 'special regard shall be paid' to the claims of schools, or the like. This will give the department some guidance, but its officials will still be left with some discretion in arriving at the exact quantitative allocation that has to be made in the end, and here they can be guided only by their own judgement of the strength of the arguments submitted and the importance of the ends to be served.

When the device of a pricing system is used to work out an allocation, these difficulties do not arise. The different claimants do not now argue against each other, they bid against each other, and the one who is prepared to pay more can bid supplies away from the one who will not go so far. The essentials here are that the allocation is made not by the administrative decision of one man or council but by the joint action of all those concerned, and each can

adjust in a continuous gradation the strength of the pressure that he exerts, so that each has some say in the outcome, and there is interplay between the strength of his demand and that of others.

It is true that if the initial distribution of purchasing power between the bidders does not seem to us right, then the allocation which the pricing system brings about will not seem right either. But equally it is true that the administrative decision may be prejudiced and unjust, and in comparing the merits of the two devices, solely as such, we can suppose that each is working in the most favourable setting. We can suppose, that is to say, that the administrative decision will be taken by an entirely public-spirited and impartial man or council, and that the pricing system is working in an economy where the distribution of purchasing power is fully adapted to personal needs and to such social needs as the maintenance of incentives.

In such conditions, the device of the pricing system has two decisive advantages over the device of the administrative decision. With the best will in the world, the administrator can only guess at the strengths of various needs and claims, whereas in the pricing system these order themselves on a common scale. The claims being entered, and an excess of demand appearing, the administrator must take the claims as they stand, and impose some cuts: but in a pricing system the excess of demand makes itself felt in a rising price, which causes each claimant to adjust his own claim, according to the intensity of his wants, and drop those parts he can best spare. Thus a pricing system can leave us free to make our own choices, while ensuring that those choices are brought into consistency with the choices of other members of society.

THE SOCIAL SIGNIFICANCE OF COSTS

In bringing about that distribution of productive resources between uses which, with any given distribution of incomes, will best satisfy the demands of consumers, the pricing system works through costs. Costs are not just part of the score card of the money-making game, part of the calculus of private profit, but are incurred, implicitly if not explicitly, and must be reckoned with, not less in the Crusoe economy and in collectivism than in private enterprise.

To achieve the 'optimum allocation' we have to be able to tell whether each parcel of resources is yielding in its present use not less than it could yield in any alternative use. A pricing system shows this through the comparison of revenue and cost. The revenue expresses the valuation which consumers place upon the present use of this parcel of resources. The cost is made up of the market prices of these resources, and expresses the valuation which consumers place on these resources in alternative uses. For if, in a certain district, land of a given quality generally commands a rent of a pound an acre, then its marginal value product in its present uses will be not less than that: taking an acre of it away from one of these uses would reduce output by an amount for which consumers would be willing to pay a pound. Then if there is a proposal to develop in the district some new line of output that would need to use this kind of land, the test of how much land, if any, should be transferred to it lies in a comparison of its marginal value product in the new use with its market rate of hire of a pound an acre. So far as it can earn more than a pound an acre in the new use, the transfer of it will be justified, for what the consumers lose in the lines of output from which it is withdrawn they value less than what it will add to the new output. It is the function of the market rate of a pound an acre to express the loss of benefit which consumers will suffer if an acre is withdrawn, and so to set the test which new uses must pass if they are to have land transferred to them.

Costs appear usually as payments of money, but if by a cost we mean a forgoing, a giving up, then a transfer of money – perhaps only pieces of paper of no use in themselves – may express a cost but is not one itself. Beneath the transfer of money lies the surrender of a claim: if I am to have this I must pay for it, because if I use productive resources to make it they cannot be used for making something else, and my giving up the power to buy something else offsets the forgoing of alternative output that is the real cost of making what I get. Costs are reckoned in money, and money may be only a unit of account, a book-keeping token, yet costs are real. Crusoe does not pay money to himself, but the cost to Crusoe of the fish he gets in a day's fishing is the going without whatever next best use he could have made of the same time, the table he could have knocked up, or the rest he could have enjoyed. A community does not pay money to itself, but the cost of its taking land for a housing estate is the going without the food that this land could otherwise have provided; and the payment of money

costs between members of the community expresses this real cost. The test for the efficient use of resources is that what each unit adds to output in its present use shall be valued not less highly by consumers than the most valued product which it could provide in another use. The rule that a firm must pay its way and cover its costs is the practical application of this test that a pricing system makes possible.

There is one other way in which the social significance of costs may be considered. The total of all money incomes is the same as the total money value of the use of all productive resources: the national income is the sum of the incomes of all persons, and also of the annual hires or earnings of all productive factors. The money income that goes to any one man, then, allocates to him the use of a certain proportion of the total resources of the economy. If he is to be able to realise this allocation of so much money as the use of some particular parcel of resources, these resources must be reduced to common terms by being priced. The rule that what he pays must cover the cost of what he buys ensures that he shall give up his claim in proportion as the resources whose use he absorbs are capable of uses valued by others. If he sets out to absorb the use of some factors which other men do not much want to use, he will be able to take in a good deal of them without thereby requiring these others to forgo what they much value: this will be shown by the cost being low here. But if the factors for which he is bidding are in strong demand by others, each unit he draws to himself will be much missed by others, and this will be shown by a high cost which prevents his money from going far. The prices which the market sets on productive factors register and balance against each other the competing pulls of consumers as they realise their money claims on the use of resources. The cost which each consumer is required to cover if he is to get a given product expresses the valuation which others set on the use they could make of the resources required to make it, and the extent therefore to which by taking it he has used up his own claim.

THREE TESTS OF EFFICIENCY

Since costs provide a test of whether resources can be applied in some proposed use to as great advantage as elsewhere, they provide a test of efficiency. If we take the maximisation of benefit to consumers as the end of activity, then the test of cost and revenue is indeed the final test, and we can bring out its meaning by contrasting it with two other tests which correspond with partial or relative criteria of efficiency.

First, there is physical productivity: for example, horse-power generated per calorific unit consumed. Here we take output, and an input, both in physical terms, and we reckon that arrangement most efficient in which output per unit of input is highest. A common use of this criterion appears in the testing of the efficiency of a firm or industry by reference to its figures of output per man-hour (*PMH*). In the same way we may compare the miles per gallon given by different cars, or the bushels per acre obtained by different farmers. There is sense in this kind of test, in so far as a low rate of output per unit of input may arise from avoidable waste, or a backward technique, and in that case a rise in the rate marks the ability to get more produce out of given aggregate resources. But the examples taken remind us that a higher rate of output per unit of input of one factor may be attained not only by changes which save factors but also by those which absorb them.

A high physical efficiency in a heat engine, for example, in the sense of a high output of power per calorific unit consumed, may be attained by elaborate insulation, which is costly to install, and raises the cost of adjustment and repairs because it makes the engine less accessible: common sense tells us that there may come a point in the raising of calorific efficiency at which the game ceases to be worth the candle, because the cost of further improvement is greater than the cost of the fuel saved. A big crop per acre, again, may be attained by heavy fertilising, and elaborate cultivation: but 'high farming is not always good farming', that is, the extra materials and labour used to get the high yield may cost more than the extra output they cause will sell for. We cannot therefore say that the arrangement which gives the highest output per unit of input of some one of the factors used is always the most desirable arrangement; nor that if the output per man-hour is higher in one firm than in another making the same product, we should always be better off if the second firm adopted the methods of the first. It is tempting to say such things, because a change which raises physical productivity is often one which increases our power to get produce from given total resources; but it may also raise output per unit of one kind of input only by using more of

other kinds. It is tempting to say such things, too, because a rise in the output per man throughout an economy is certainly a gain for it; but a rise in the output per man in one sector of the economy may be brought about by absorbing factors that would have been of more use elsewhere. (The presumption that what holds of the aggregate also holds of the part is the inverse of the usual fallacy of composition. It specially needs watching in international comparisons of output per man-hour in a given industry.)

We recognise, then, that output per unit of some one input cannot provide a reliable test of efficiency, if by efficiency we mean 'getting the greatest benefit from given resources'. But physical productivity can be reckoned only as output per unit of one kind of input, for we cannot add different kinds of input together, in their physical terms: so it remains at the most a provisional or partial test. Evidently the final test must weigh not one alone but all of the factors used, against the product made. When a change of method is proposed, the test must be whether the extra use of some factors that it will require will be more than offset by the savings it will make in the use of others. But this is a balance that cannot be struck in physical terms. We cannot say whether the use of so many more kilowatt-hours will be offset by the saving of so many man-hours. 'Obviously,' we should say, 'this depends on how much the power and the labour cost, and the change might be worth making if the power were cheap but not if it were dear.' In saying this we are recognising that factors can be weighed against one another only through valuations. When the factors bear market prices derived from consumers' valuations, the use of these prices is a device which both solves the problem of entering up amounts of different factors in a common unit of account, and provides in the total cost so computed a measure of the aggregate significance of all the factors absorbed.

This brings us to the second partial or relative test of efficiency, namely that which asks whether a certain output is being obtained for the smallest possible total cost. Whereas in the first test we looked only at the output per unit of some one factor, we now take all the factors, and add them together by means of the common unit of account which the pricing system gives us, and we look at the output per unit of this inclusive cost. The method of output which in this test emerges as the most efficient will depend not on technical issues alone but on prices, and it will change as prices change. At the prices of today the most efficient engine may be one which is elaborated so as to conserve heat, but if the price of fuel falls relatively to that of machinery, the most efficient engine will now be one of simpler construction that loses more heat. In a country such as the prairie states of Canada, where land is plentiful relatively to labour, the most efficient method of growing wheat will show a low yield per acre: in England, where these relations are reversed, it will show a high one. For any proposed change of method the test is whether the cost of the additional factors which it will require, to get a given output, will be less than the cost of those it will save.

This test is not liable to mislead as the first test was, for it cannot show an apparent rise in productivity when in fact the total absorption of resources per unit of output is greater than before. Yet it still is partial, or relative, for in itself it cannot tell us whether the resources absorbed in a given line of output ought to be used there or somewhere else. Given that a certain output is to be made, then it will show which of any two methods is the better way of making it, and so arrive at the best method of all: it does this by calculating a total cost which expresses the total draft which the given output is making on the resources of the economy, and indicating that method as the best which makes the smallest draft. But when this has been done, we still do not know whether it is an efficient use of the resources of the economy to make this particular output at all: that is, whether the benefit justifies the draft. By the same token, though this second test can decide whether or not a proposed change in the method of making a given output should be made, it cannot decide whether a proposed change in the size of that output should be made, by whatever method it is to be effected. For to answer such questions as these, we must compare the cost with the gain. We are now asking, not how best to make a given output, but how best to deploy our available resources between the making of different kinds of output, all of which have value to consumers.

If our purpose is to maximise the economic welfare of consumers, the test which answers this question is the third test of efficiency, and the only complete one. It is the pricing system that provides it.

SOME QUALIFICATIONS

Our exposition of the functions of a pricing system and of the nature of welfare economics has not provided an account of the working of any actual economy. Nor has it provided a justification of any system. For one thing, it has studied the workings of the pricing system only in abstraction from much else that in practice is knit up with them. It has considered that system only as a means of reckoning costs and so providing the data for the optimum allocation of resources. But the actual working of a market in which prices are settled provides also a system of incentives, a canalisation of the motive of gain-seeking, so that it does not only show what arrangements are desirable, but also gives men a motive to make them; and this function of motivation we have not considered. When, moreover, we set out the conditions for the optimum, we did not imply that these were fully realised in any actual economy: on the contrary, the absence of controls and direction is no assurance of a perfect market; imperfection of competition is sometimes inherent in the nature of things, and even were that not so, men have a propensity to combine as well as to compete; and, beyond that, there are custom and inertia to obstruct and delay readjustments. Even, moreover, did every market register the play of supply and demand as quickly and exactly as a stock exchange, we should still not be satisfied with the result if we believed that the moneys which consumers had to spend were not distributed in a way acceptable to the social sense of what is practicable and fair – and whether this is so depends on much more in the social structure than the existence of such devices as the pricing system.

The persistence of imperfections, to which we have just alluded, perplexes the choice of policy. For even if our sole interest is in the efficiency with which resources are distributed between uses, we still cannot simply infer that policy should remove all the imperfections it can. If there is only one market in the economy where the determination of prices and factor-uses is obstructed by imperfections, then welfare economics does suggest that policy should remove them, and so shift the pattern of factor-use there (and, necessarily, in other markets, too) towards the 'optimum'. But suppose that there are two such imperfect markets, and that there is no possibility of removing the imperfections in one of them: what should we do now? The 'optimum' is unattainable, and there can be no general presumption that we should get nearer to it by removing the imperfections in the market that policy can take effect on, for the effects of these imperfections on factor-use elsewhere may be serving to offset those of the irremovable imperfections.

It is a further qualification of our treatment that it has assumed the sole object of allocating resources between different uses to be the greatest aggregate benefit of the consumers when each is left to lay out his money according to his own notions of what will benefit him. In practice, societies seek other ends as well as this. We accept the need to club together to pay for objects like defence which in the nature of things we cannot buy in penny packets, each for himself. This might seem a difference of means rather than of end, but in the political decisions on how much defence to provide and how to raise the funds for it, the judgements made are of a different kind in part from those of the personal shopper: we accept the need to look ahead, and to provide for others as well as for ourselves, as few of us would do if left to spend on our own. Thus we recognise that in the protection and progress of our society there are ends we acknowledge no less than the end of personal benefit from consumption.

It is true that some of these other ends can be and often are sought through the pricing system, in which they compete with the personal spending of consumers for the use of resources: it being decided, for example, as a matter of public policy, to build a certain school, the public authority can take money away from consumers in taxes and use it to bid in the factors needed to do the building, and the pricing system will respond to the spending that results from such a collective decision in the same way as it does to the spending that each man decides on for himself. But there are other ends that cannot be attained simply by a change in the pattern of final demand: it may be decided, for instance, to call men up for national service, to forbid the employment of women at night, or to give a preference to imports from a particular region.

The example of the employment of women at night prompts the reflection that the ends we acknowledge include benefit in the working life as well as in consumption. The consumer is only part of the whole man, who is also a worker, and to most of us the

satisfactions afforded by work matter no less than those consumption yields. Once again, the pricing system will take care of this in part. We 'spend our time' in some ways much as we spend our money; and where workers of a given type have an effective choice of jobs, they will get paid more for the jobs they find less agreeable. But considerations of foresight, of effects on the growth of society, and of justice between persons, enter into our collective judgements about the spending of time no less than of the spending of money: we may protect agriculture, for example, because we believe it fosters a valuable type of man and home. The pricing system provides for a double adjustment, of the work men get to the products they want, and of the products they get to the work they want; but the second adjustment calls for some collective intervention no less than the first.

For all these reasons, our account of the working of the pricing system, and its bearing on the allocation of resources, must be recognised as simplified and abstract. Yet the fact that it has provided only a view of the functioning of one social device, in abstraction from any actual setting, and has suggested a criterion of only limited value by which to judge the outcome, does not mean that we have discovered nothing of value. For however important the other ends of human activity may be, the fact remains that consumer's benefit is a main object of the day's work in any society outside the beehive and the ant-hill. A calculus that will show how to use resources effectively for this object, while it is not conclusive for the choice of policy, has a value that is unmistakable. Though, moreover, our argument does not provide a vindication of private enterprise, it does pose a problem for the organisation of a collectivist society. For the functions we have isolated do have to be discharged in any economy in which there is division of labour. The markets of a private enterprise economy discharge these functions imperfectly, but they do discharge them somehow. The question is how in a collectivist society they can be discharged at all. In trying to answer this question [one may] throw further light on the problems of resource allocation.

Notes

1. Defined as the change in production (measured in terms of money by multiplying the volume of output by the selling price of the product) which would occur if the input of a given factor of production were reduced by one unit; e.g. if the result of dismissing one worker were to reduce output by £25 worth of goods per week, labour's 'marginal value product' in that line of employment would be £25 (Ed.).

2. 'Cost' is here defined as including the 'normal' profit which is just sufficient to induce the owners of the firm to undertake the business and remain in it. 'Entrepreneurship' – the bringing together of labour and capital in a productive enterprise – is thus regarded as an additional factor of production, whose remuneration is included in 'cost' (Ed.).

THE SOVIET ECONOMY: MICRO-ECONOMIC PROBLEMS

Alec Nove

'SUCCESS INDICATORS'

In the West, firms take decisions by reference to the profit-motive; in the U.S.S.R., all decisions must be based on and conform to the plan. This generalisation, while not wholly accurate,[1] does underline an essential difference between the systems, despite the fact that considerations other than direct profit expectations do influence actions of managers in the West, and profits do affect to some extent the behaviour of managers in the U.S.S.R. The primary task of a Soviet manager, as we have already noted, is to fulfil the plan, or more strictly plans, since he is judged and rewarded under a number of plan indicators. We shall now consider the effect of this on Soviet managerial behaviour and on the output of goods and services, beginning with the influence of these 'success indicators' on the quantity and assortment of output.

This influence is related to the range of choice open to managers at enterprise level, and this in turn depends on two factors: the nature of the product and the degree of detail in which the plan is elaborated by higher authority. Thus, for example, an enterprise producing electricity, or a factory making large specialised machines (e.g. turbines), or arms factories producing to the precise specifications decided centrally by military experts, whatever their other differences, have in common the fact that the product mix depends hardly at all on decisions taken by enterprise managers. In most cases, however, the problem of the product mix is a real one: textiles, shoes, a wide range of metal goods and a long list of other commodities can be made in many shapes, sizes, styles, qualities, colours. It is generally impossible for the planning organs to specify

the desired assortment in full detail, or, as we shall see, to enforce its specifications on the enterprise even if they made the attempt to subdivide the product mix. It is also important to note that the actual or potential production of any factory must always depend to some extent on the management on the spot, since the plan itself can hardly fail to be influenced by what the management declares itself able and willing to do in the next planning period.

The system is now in process of transition. In analysing its problems, we must never lose sight of the fact that many enterprises are still not 'reformed', and that, as will be shown, the reform is only partially operative even where it is said to apply. So this chapter can and should describe the problems of the 'traditional' planning system which may now be withering away, as well as the growing-pains of the new model which is taking its place.

It was the essence of the traditional system to specify plan indicators from above in some detail, and then to reward fulfilment and overfulfilment of these plans. If the fulfilment or overfulfilment of plans is to be rewarded, it is clearly necessary first to define what the plans are. To take three of the most common success indicators as examples: in each period there must be specified a quantity of output, a percentage reduction in costs, and a percentage increase in labour productivity. Managerial behaviour must naturally be influenced in the direction of endeavouring to fulfil the plans so defined, with results which may not be obvious at first sight.

Take the quantity of output first. The plan was generally based on the achievements of the previous period, to which was added a percentage increase. The quantity to be produced was usually intended to encourage effort, to induce the management to seek out 'reserves' of under-used factors of production. In the process of plan determination, it pays the management of the enterprise to conceal its full potential and in other ways to manœuvre to obtain an 'easy' plan; nor would it pay a wise manager to overfulfil it by too wide a margin, for, if he did so, he might be suspected of having concealed his potential, and plans for the subsequent period might be increased by an embarrassingly large percentage.[2] Of course, all this is known to the planning organs, and in the resultant manœuvring the authorities try to utilise the various supervisory agencies, to check up on managerial behaviour. Plans may be deliberately 'tightened', on the not unreasonable assumption that the manager is not being quite honest about his production possibilities or his stocks of materials (as we shall see, it paid to hoard). These circumstances also help to explain the urge to 'overfulfil' output plans, which perfectionist critics of planning often regard as proof of inefficiency, but which springs in part from a knowledge that there are hidden reserves. On the other hand, the 'supervisors' are often themselves interested in being able to report plan fulfilment. Thus ministerial officials tolerated understatements of production possibilities, while a *sovnarkhoz* was said to have reduced a plan in mid-year so as not to have to report that one of its major enterprises had failed to fulfil it.[3]

However, let us now assume that an output target for the next plan period has been decided. One must now consider how the target figure is expressed in quantitative terms. The ways in which this can be done depend in part on the nature of the product: for example, shoe output could be expressed in number of pairs or in roubles, production of clothing or of furniture is too varied for the plans to be expressed in anything but roubles, sheet metal plans could be in tons or in square metres, cloth in square metres or linear metres or roubles, and so on through a variety of other possibilities. The effect of each of these is to encourage its own species of distortion. In the context of an economic situation in which virtually all goods free of major technical defects can be disposed of, the temptation always exists to adapt the assortment of the product to the method of measuring the quantitative fulfilment of the plan. It will be noted that the substitution, after 1965, of *sales* for gross output cannot solve the problem, so long as there is a sellers' market.

Examples can be cited in very large numbers, drawing on material published not only in the U.S.S.R. but also Hungary, Poland and Czechoslovakia, where a similar institutional structure involved similar problems.[4] A plan expressed in tons encourages the production of heavy commodities; in any choice involving weight, the 'weightier' variant is bound to be favoured, since this facilitates the fulfilment of the plan. Thus, for example, a metal works was reported to have increased its output of roofing iron by 20 per cent over five years in tons, but in terms of area the increase was only 10 per cent; the plan, of course, was in tons.[5] Clearly, an output plan expressed in terms of tons encourages the choice of a variant in which a given weight is achieved with the least expenditure of those resources unrelated to weight. For instance, factories

making prefabricated cement blocks prefer to make large blocks, which is the easiest means of fulfilling a plan in terms of tons, though, as it happens, the result is a shortage of small blocks necessary for completing portions of the buildings under construction.[6] Khrushchev himself gave a particularly absurd example: the chandelier plan was expressed in tons, so chandeliers were unnecessarily heavy.[7] The humorous journal *Krokodil* once pictured, in a cartoon, a factory which fulfilled its entire month's output programme for nails by the manufacture of one gigantic nail, hanging from an overhead crane the whole length of the workshop.

Output targets in roubles evade such difficulties as these, but at the cost of creating others. In most instances, the money measure was applied to gross value of output – *valovaya produktsiya* – which includes unfinished production, or, strictly speaking, the increase (if any) of unfinished or partly-processed goods over the previous plan period. This encouraged a number of distortions. In the first place, an advantage is derived from using dear materials. Obviously, to take a published example, an enterprise making tools, with a plan expressed in roubles of gross output, finds that it 'pays' (in terms of plan fulfilment) to use unnecessarily expensive quality steel, and for many years the production of inexpensive clothes was inhibited by the reluctance of manufacturers to utilise the cheaper cloths.[8] There is even involved in this method of measuring plan fulfilment a quite unintended discouragement of the production of spare parts. This is because, for example, a carburettor is 'worth' more in a completed motor-cycle or tractor, where it is combined with other goods and services many of which were bought from outside the enterprise, than if this same carburettor were produced and sold as a spare part.[9] Spare parts are notoriously short in the U.S.S.R., and it is noteworthy that a plan expressed in physical units also leads to neglect of spares. Thus *Pravda* (8 June 1960) published a bitter attack on those who fail to provide spares for motor-cycles, in the course of which an imaginary person, deemed to be 'the director of an enterprise, head of a *sovnarkhoz* or official of a planning organ', undertakes to increase production of motor-cycles, and, when the question of spares is raised, answers as follows: 'I am not interested in your calculations. All that matters is what is included in the statistical returns and counted in the plan, i.e. new complete machines. They interest me. As for piston rings, plugs and so on, these are just fiddling little things which don't appear in fulfilment indicators.' Spares, therefore, are unobtainable, for reasons connected with the success indicator system. At the same time, the failure of British car manufacturers to provide adequate spares and service in the first post-war boom years should remind us that these things are apt to suffer in a seller's market, even in a quite different institutional setting. . . .

Other measures of output plan fulfilment suffer from defects of their own. Cloth was 'measured' in linear metres for many years, with the result that it was made narrower than was desirable from other points of view.[10] Consequently from 1959 the basis was changed to square metres. But this is by no means always satisfactory, since other desiderata tend to be sacrificed to maximise square metres, at the expense, for instance, of quality or workmanship. Roofing metal which was too heavy when the plan was in tons becomes too thin when the plan is in square metres. If a plan for nails were expressed in quantity (e.g. thousands of nails) they would tend to be small, if in tons they would tend to be large. Highly original plan-measurement criteria were devised in some industries; for example, central-heating boilers were assessed for this purpose in terms of the area of heating surface (of the boiler); consequently, when a new model was devised which heated more efficiently with a smaller heating surface, no one would touch it, as it would worsen their success indicators.[11] The Soviet economist Aboltin reported that the Stalingrad construction trust wasted metal because, for one of its projects, the plan was expressed in terms of metal used. In the course of an admirable survey of many of these distortions, the aircraft engineer O. Antonov cited the case of the management of a road-transport enterprise which ordered useless journeys, often with empty lorries, at the end of a plan period in order to record the planned number of ton-kilometres. . . .

Rewards for plan fulfilment require not only a clearly defined measure of results, but also a time-period to which the results must be related. Hence the great importance of the calendar, and the phenomenon of 'storming' (*shturmovshchina*), of a mad rush to fulfil the plan in the last few days of the month, quarter or year, followed by a slack and disorganised period in which production falls sharply until the next mad rush. Denunciations of *shtur-*

movshchina were, and are, extremely common, and the prevalence of the disease is a sign that it is, in a sense, built into the structure of the economy. It is worth recalling that this particular form of distortion, as well as many of the other 'success indicator' problems referred to here, occurred also in British planning during the war.[12]

One weakness involved in 'calendar' planning is a certain short-sightedness; one has insufficient incentive to look to subsequent planning periods, the more so because, with a high turnover of directors, someone else might then be in charge. There is little or no incentive for looking far ahead. This affects attitudes to resource conservation, as may be illustrated by a large number of articles pleading, apparently in vain, for preserving forests or the avoidance of pollution of rivers or lakes.

The effect of other 'success indicators' on the behaviour of managements may be partly to correct some of the deviations engendered by pursuit of quantitative plan fulfilment, but often they encourage similar distortions, or cause new ones. Thus an effort to fulfil a plan to increase labour productivity by 4 per cent would, clearly, discourage the taking on of large numbers of additional workers. On the other hand, the calculation of productivity necessarily requires comparing the labour force to total output, and this reinforces the tendency to try to simulate overfulfilment of the output plan in tons, roubles or whatever the measure may be. Until 1958, the measurement of labour productivity involved only 'productive workers', i.e. excluded clerks, technical personnel and auxiliaries. The effect, not surprisingly, was that the management either neglected to mechanise the auxiliary work or endeavoured to classify as many of their employees as possible in non-productive categories. This particular deviation has been checked; from 1959, labour productivity statistics are based on the entire labour force.[13] . . .

It is essential to appreciate the cause of these and similar difficulties. They all spring from the fact that the activity of enterprises and subordinate planning bodies is essentially based on *instructions*, and the incentives provided to the management are rewards for acting in accordance with instructions, with the plan.[14] The authorities are unable to issue unambiguous, all-inclusive instructions, and various aspects of the plan may be inconsistent with one another, as when overfulfilment of output plans may require some increase in costs, or when some other plan indicator or instruction (e.g. relating to saving fuel, or economising in the use of some scarce material) contradicts the aim of increasing labour productivity, or involves the overspending of the wages fund. These ambiguities are further complicated by the inevitable imperfections and imprecision of the statistics which express the achievements of the enterprise by reference to the various indicators. . . .

The relative failure to adjust production and distribution to the requirements of the consumer has three aspects, or causes, which ought not to be confused in analysing defects or in endeavouring to see ways and means of overcoming them. The first, which has been much emphasised in the last few pages, is the unresponsiveness of the price system to demand, and, on the contrary, the responsiveness of directors to 'success indicators' and plans which cannot of their nature take consumer demand adequately into account. The second aspect concerns the habit of underpricing, the tendency towards tolerating queues rather than raising prices to the extent necessary to eliminate them, with the consequence of a chronic seller's market, with all that this implies in terms of quality of service and, indeed, responsiveness to demand.[15] The third is a question of priorities: during most of the Soviet period, scarce human skills and material resources have been systematically used in heavy industry rather than in consumers' goods industries or in trade; in these non-priority sectors, investments have been relatively low, the quality of managerial personnel relatively inferior. In the case of retail trade in particular, the effect of 'success indicators' is often extremely misleading. Thus labour productivity (in terms of turnover per shop assistant) is higher and distribution costs are doubtless far lower in the U.S.S.R. than in Great Britain, but this is in large part due to such factors as the far smaller number of shops per thousand customers in the U.S.S.R., smaller stocks in relation to sales, much more intensive work of the shop assistants. Indeed, most British shop assistants spend a large part of the day waiting for customers, so in a sense they are 'inefficiently' used, though this is the inevitable consequence of a wider degree of consumer choice and the absence of queues.

But having said all this, it is important for the student of Soviet economics not to overlook a very essential point: whatever may be the inefficiency of the mechanism by which goods are distributed,

whatever the inadequacies of the finer adjustments to the citizen's requirements, more goods tend to be produced, retail trade turnover increases by impressive percentages. In certain circumstances, even the appearance of a long queue may represent progress; there is no queue for goods which are not there. The Soviet system does, in normal years, ensure that there will be many more raincoats in three years' time than there are now, even though the styles, sizes, quality and range may all leave much to be desired. More cloth, clothing, shoes, television sets, handbags and so on are being produced. In so far as this is so, a steady increase in living standards does occur, and Soviet economists have some reason on their side when they accuse their Western *confrères* of being so concerned with optimum distribution and the finer adjustments of output to demand that they overlook the quite essential point of the quantity of goods produced. A million raincoats, though they may not precisely correspond to what the citizen desires and may have to be queued for, are surely better than half the number produced and sold under the most perfect principles of the market and the capitalist price-mechanism.

This argument is certainly not to be ignored. Yet it contains a flaw which modifies its effectiveness considerably. Certainly, *ceteris paribus*, a million raincoats are twice as numerous as half a million raincoats (if one abstracts from the fact that other goods may be more urgently required). However, in terms of use-value, it is not always so. Two raincoats which do not fit are not worth twice as much as one which does. A given number of raincoats in fully assorted sizes and styles, which correspond to consumer requirements, are not 'equal' to the same number of raincoats which are available only in two styles and two sizes, in any but a purely numerical-statistical sense. All this is in the highest degree relevant when comparing Soviet output of consumers' goods with that of a 'capitalist' country. Standardisation and a restricted range of choice are not necessarily a bad thing, but they have certain consequences in terms of aggregates of consumer satisfaction, which are none the less real for not being measurable. If a Soviet housewife wishes to buy a certain article and has to make do with another, her annoyance cannot be measured; only the thing she actually buys is statistically recorded. Yet in our conventional measurement of 'welfare', or in our aggregation of the value of goods, it is implicitly assumed that the consumer can choose from a wide range and that these goods are available at established prices. It is a scarcely deniable fact that the Soviet economy falls short in these respects of Western standards. Plan-orientated production choices in the U.S.S.R. are less effective instruments for adapting production to demand than are profit-orientated production choices under 'capitalism', sceptical as one may be about the reality of so-called consumers' sovereignty in the Western world. This must have some unmeasured and probably unmeasurable bearing on the relative value of the consumer goods produced under the two systems.[16] As will be shown subsequently, all this has a bearing also on growth rates.

The reforms now being implemented have as one of their principal objects the correction of these distortions. To some extent they must be disappearing. Thus *sovkhozy*, especially those who are being granted greater financial autonomy, are empowered to refuse equipment which they do not need. It is also intended that contracts between enterprises, albeit under the 'umbrella' of allocation decisions by Gosplan and *Gossnab* (and sometimes via freely negotiated direct links), should more fully express the requirements of the customer-enterprises. The emphasis on profitability and *sales* as the principal success indicators undoubtedly detracts from the 'gross-output-for-output's-sake' approach. No sale, no bonus. In so far as the enterprise's own proposals affect the plan finally adopted – and to some extent this has been so even in Stalin's time – the new success indicators do orientate managerial thinking towards a study of demand, of the market. This applies also to consumers' goods. Trade organs have been urged to study demand, have been given greater flexibility. It has been recognised that one of the principal causes of the failings of retail trade has been the illogical and sometimes plainly absurd financial strait-jacket within which they operated, and retail trade enterprises are being given some greater leeway, with fewer indicators imposed from above. However, retail margins are often much too small, handling and storage facilities inadequate.[17] No doubt all this is historically explained by the prevalence of shortage, of 'goods famine', for a generation. Yet now that overproduction is by no means uncommon, the new methods now being introduced should transform the situation, according to optimists.

But will they? Some of the causes of past troubles are still with

us. Sales are perhaps a better indicator than gross output, but it is still a gross-value indicator. Material supplies are still, in the main, administratively allocated. The principal output targets in physical terms are still planned from above. Above all, prices are still inflexible, and profit margins are fixed (as in wholesale and retail trade), and are unrelated to fluctuations of demand, to scarcity or to abundance. . . .

Notes

1. It does not apply, for instance, to most nationalised industries in the West.
2. These defects are discussed by G. Nikolaeva, in *Vovyi mir*, no. 7/1957, p. 74, comments on readers' letters in *Kommunist*, no. 1/1957, p. 49, and many others.
3. M. Sidorov, *Vop ekon.*, no. 4/1957, p. 128, and *Pravda*, 27 April, 1959.
4. Particularly valuable in all matters pertaining to the present chapter is the book by the Hungarian economist Janoš Kornai, translated into English under the title of *Overcentralisation in Economic Administration* (Oxford, 1959). For a discussion of Soviet experience, see A. Nove, 'Some Problems of "Success Indicators" in Soviet Industry', in *Economica*, January 1958, pp. 1 ff.
5. V. Kontorovich, *Sotsialisticheskii trud*, no. 1/1957, p. 50.
6. But if the cement blocks are calculated in roubles, they avoid making cheap cement blocks (*Pravda*, 3 August 1956).
7. *Pravda*, 2 July 1959.
8. See Kontorovich, op. cit., and N. Lyubimov and A. Petrov in *Plan. khoz*, no. 1/1957, pp. 38–9.
9. This point is made by N. Antonov, *Plan. khoz.*, no. 5/1957, p. 83, and V. Kontorovich, in *Plan. khoz.*, no. 3/1957.
10. Kontorovich, op. cit., p. 50.
11. *Pravda*, 5 September 1958.
12. See, in particular, E. Devons, *Planning in Practice* (Cambridge University Press, 1950).
13. A. Rumyantsev, (Ed.), *Ekonomika Sotsialisticheskikh Promyshlennykh Predtriyath*, Moscow, 1959, p. 271.
14. This thought is developed at some length, with impeccable logic, in the excellent book by Kornai, cited in note 4 above.
15. Note the interaction of these two aspects: it is largely because of queues – i.e. of underpricing – that retailers, wholesalers and manufacturers can adopt a take-it-or-leave-it attitude, and concentrate on success indicators' without fear of not selling whatever is produced.
16. The following somewhat extreme example of 'output' for plan fulfilment instead of for profit might make the general point clearer to some readers. Suppose that a Soviet publisher of popular songs has a plan of 100 songs; a British pop-song publisher has no plan, but seeks to maximise profits. The Soviet publisher, other things being equal, would tend to publish more songs (at least 100), because, until he fulfils his plan, he would choose to publish in all cases of doubt. Yet, even on the assumption that an increase in output of pop songs is a good thing (which God forbid!), what meaning would this 'superiority' have in any conceivable terms of worth, value of use? For a rigorous statement of similar arguments, see M. Polanyi, 'Towards a Theory of Conspicuous Production', *Soviet Survey* (October–December 1960).
17. See, for instance, *Sovetskaya torgovlya*, no. 10/1966, pp. 6 ff., and V. Mineev *et al.*, *Vop. ekon.*, no. 4/1967, pp. 39 ff.

Each of these pieces is a clear and simple exposition of a very different theme. Each, however, shows how deeply embedded in social *structures are the workings of economic systems; the history, values, attitudes and interpersonal behaviour of men and women play a large part in shaping economic activities, while much social behaviour in turn is shaped or influenced by the operation of the economy.*

The first piece about labour–management relations in Japan was written by a Japanese social scientist during the mid-1960s, a period during which the Japanese rate of economic growth was outstripping that of the leading Western economies. As this article reveals, a different set of social values and relationships have resulted in Japan in a type of labour–management relationship radically different from that in Western Europe or the U.S.A. This is a reminder that there is nothing inevitable about the type of 'trade union', for instance, that we have in Britain.

The second article by Roy first appeared in the 'American Journal of Sociology' in 1955. This fascinating and highly amusing article shows some of the ways in which a group of workers seek to control the work situation in face of a management-inspired piecework system of payment. Such 'informal' behaviour is, of course, known about by management.

LABOUR–MANAGEMENT RELATIONS IN JAPAN

Shizuo Matsushima

I

In any discussion of Japan's trade unions, great emphasis is always laid on the fact that the overwhelming majority are organised as 'enterprise' unions, that is with the enterprise or place of work as the basic unit of organisation. The Basic Survey of Labour Unions, Japan's most authoritative trade union survey, shows that up to 90 per cent of trade unions are organised on this basis, membership consisting of *all* employees in a given establishment. Unions organised on a national or craft basis, which is the commonest form in other countries, are very rarely found in Japan.

It is inevitable that this characteristic should have had much influence on trade union activities in Japan. Since members do not join a union of their own choice, the unions will always contain a certain proportion of workers whose 'trade union consciousness' is undeveloped; thus, if union leaders attempt any very militant action, there is always the danger that members will defect and split the union. Further, since the union is based on the enterprise, both management and union leaders belong to the same firm; in these circumstances, the firm's profitability is bound to be a determining factor in any discussion or negotiations. Union activity has therefore tended to be geared to the pace set by management, for management is not obliged to regard wage levels as a constant when making managerial decisions. This has given rise to wide differences in working conditions and wage levels between different enterprises and unions, and such differences make it even more difficult to carry on concerted negotiations on an industrial or craft union basis.

Attention has often been drawn in Japan to the importance of strengthening the national unions as a means of overcoming this weakness of the Japanese trade union movement, but efforts in this direction have borne little fruit, and the enterprise unions have remained dominant. The national unions thus give an impression of being little more than loose federations of enterprise organisations, serving mainly as political groupings. The rules of the national unions do not preclude the admission of individual members, but it is normal practice for the enterprise union to be considered as the unit both in the case of affiliation to and withdrawal from the national body. In addition, when the national union is involved in a dispute, the final right to agree to any settlement is usually retained by the enterprise unions; there is very little uniformity of opinion among them, so that the national unions must spend a great deal of effort in controlling their constituent enterprise unions.

As a result, although the enterprise unions have been quite ready to obey directives to strike on political issues – the most radical type of union activity – they are very reluctant to take any decisive action on issues likely to involve friction with the manage-

ment of their own enterprise – issues which, in a sense, constitute a trade union's proper field of activity. Japanese trade union activity has thus taken on a distinctive character – extremely left-wing in ideology, but right-wing in the matter of action; this latter tends to take place within lines laid down by the management, except where political issues are involved.

II

What, then, are the factors which have given rise to this peculiarity of Japanese trade union activity? The first point we should keep in mind is that the pattern of personal management – the policy of the enterprise towards its workers – is very different from that found in the majority of other countries; in Japan, very great stress is laid on the security of the workers' livelihood.

This attitude is revealed first of all in the engagement of new employees by the company. Apart from candidates for executive posts, which are publicly advertised and for which a fairly stiff examination has to be passed, a very large number of firms engage employees on the strength of family ties or some other close relationship with men already employed by them. Very many firms have found no difficulty in filling their vacancies from applications received in this way, without any recourse to recruitment on the open market. This method of recruitment, of course, narrows the field of choice and precludes the possibility of selection from a wide range of capable applicants; consequently, there has recently been an increase in the number of firms which, in order to keep up with the pace of technical innovation, have felt the need for workers with a high level of basic knowledge and have therefore abandoned this practice in favour of recruitment by public advertisement. Nevertheless, the method of recruiting through personal connections has the overwhelming advantage that accurate information can be obtained about the applicant's ideology, attitudes and skills, and that it is an easy way of obtaining workers likely to identify themselves with the company's interests. The vast majority of Japanese employers, if asked whether they preferred loyal but inefficient workers to those who were efficient but had no sense of loyalty to the company, would choose the former. For the Japanese firm, placing as it does so much emphasis on loyalty, this method of recruitment through personal connections has been ideal. The majority of employees recruited in this way have been the children or close relatives of men already working for the firm. Since the worker can thus in many cases continue to live in accommodation provided by the company and to use company welfare facilities, even after his retirement, this practice has served to give the worker a substantial degree of security, for, with living costs at their present level, retirement benefits are not sufficient to assure a livelihood to the retired worker.

The characteristic attitude of the Japanese employer to his employees can also be seen in the *terms of employment*. This phrase is normally taken to refer to the purchase by the employer of a certain amount of labour for the length of time he requires it, in return for which he pays wages. In Japan, however, though the relationship is superficially the same, it does not in fact consist in the purchase of labour *only for the length of time such labour is required*. The distinctive feature of the employment relationship in Japan is that a firm (and particularly a large firm) makes a practice of recruiting annually from among those who have left middle school, high school or university in that year, and these new recruits will take up employment with the full intention of staying with that same firm until retirement age. Once a company has employed a worker it will not attempt to dismiss him unless circumstances make dismissal quite unavoidable. It would seem, as Abegglen pointed out in *The Japanese Factory: Aspects of its Social Organisation*, that once a worker has been taken on, the firm relinquishes the right to regard him at any subsequent time as unsuitable for the post. For example, in a factory employing 4350 workers discussed by Abegglen, he was astonished to find that only five or six men were dismissed each year, and this only when no possible mitigating circumstances remained – cases of persistent absenteeism, or of men who had not appeared at the factory for several weeks.

The relationship between the Japanese firm and its employees is thus a permanent one. This has meant that, compared with other countries, the movement of labour between firms in Japan has been very small and that firms have come to regard their labour costs almost as a component of their fixed costs. Consequently, there has been a lack of the flexibility and elasticity in the labour force that a firm needs to cope with fluctuations in business activity. In order to overcome this weakness and provide some flexibility in the work-force, there has grown the practice (rarely seen in other

countries) of employing a section of the workers on a temporary basis, with short contracts of employment which can be renewed at intervals of from one to three months. In addition, since the labour market is closed, with workers once dismissed finding it difficult to obtain new employment in a large company, it is usual for trade unions to fight bitterly against any reduction in the labour force. This resistance serves to strengthen even further the distinctive character of the employment relationship.

The third peculiar feature of the Japanese firm is its wage structure. Typically, wages in Japan include large differentials based on the age and length of service of the worker, while insufficient attention is paid to the quality and quantity of the work itself. When a new worker enters a company, after a short period during which he is becoming accustomed to the work, his efficiency will rise sharply. Then, having reached a certain level, it will level out; then again, with the passage of time, it tends to decrease. Wages, on the other hand, will continue to rise in proportion to the worker's length of service, bearing no relationship to this efficiency curve. This type of wage curve is a distinctive feature of the Japanese wage system.

Payment of wages on an hourly basis is far more common than the piecework system, and the hourly wage is usually a so-called composite wage, with the worker's education, skill, length of service, age and work record all taken into consideration. Of these factors, length of service is given by far the greatest weight. This system of wage calculation, of course, has its advantages, reducing labour turnover and consequently the costs involved in recruiting and training new workers. But by far its most important function lies in the security it offers the worker. Since a worker's living costs are bound to increase as he grows older and his family becomes larger, given a low level of wages, it was inevitable that the wage structure should take this form. But, provided a worker remains permanently with the same employer, his age and length of service will rise together, so that wage increases given in return for his increased contribution to the enterprise (measured in length of service) do in fact parallel the increase in his living costs as he grows older.

Retirement benefit, which occurs far more widely and carries far more weight in Japan than in Europe or the United States, has largely the same character as wages. But not only is an individual's retirement benefit based on his final wage level (where the influence of length of service is already large), but this base figure is then multiplied by the number of years he has worked for the firm, and sometimes yet again by some other factor, depending on the company. The weight given to length of service in the calculation of retirement benefit, therefore, is enormous, even when compared with its part in the calculation of wages. It is very clear, therefore, that wages in Japan are paid in such a way as to make it quite contrary to the worker's interests to leave his employment before he reaches retirement age; this situation reflects the wish of the firm to provide security for the workers it has nurtured and trained.

The fourth factor to reveal the attitude of the Japanese firm to its employees is the provision of welfare services. Of course, to the worker, his wages are by far the biggest factor influencing the security of his daily life, but if the firm provides him with other assistance, in his role as a consumer, the real value of his wages will be increased. In Japan, expenditure by companies on this type of welfare benefit is very high. According to a survey made in 1961 by the Japan Employers' Federation, the highest figure for this type of expenditure was in the construction industry, at 8584 yen per worker per month, followed by mining at 7909 yen and the ceramic industry at 6890 yen. The average for all firms included in the survey was 4554 yen per month, while the average wage was 39,041 yen. Expenditure on welfare benefits thus amounted to about 13 per cent of the monthly wage bill.

These welfare benefits most commonly take the form of grants of money on such occasions as marriage, childbirth, illness, death and so on, and the provision of workers' housing, subsidised shopping facilities, and grants or loans in support of various cultural activities. The very high ratio of welfare benefits paid voluntarily to those which Japanese companies are legally obliged to pay reveals the extent of their concern for their workers' welfare and contrasts sharply with the practice in other countries. Of course, the provision of workers' housing near to the place of work in industries such as mining and construction, which often operate in isolated districts, and in industries where technical progress has brought with it a three-shift system of working, could be said to be a necessary development towards the provision of security for the worker; this housing should thus be regarded as part of the

productive facilities of the enterprise. But the present situation, however we look at it, is one of over-service to the worker, and even if we accept the argument that this represents an addition of so many thousand yen to the worker's wage-packet, there is a real problem in the fact that this is not paid over to the worker in cash.

The Japanese worker has always set great store by such fringe benefits received over and above his wages, which have served to cover up Japan's low wages to a certain extent and helped to bring about the companies' attitude of concern and responsibility for the daily life of the worker.

In addition to the examples mentioned above, we could mention such practices as share-ownership schemes, and various special training schemes operated by companies for their employees.

III

The term 'familism' has often been used to describe this characteristic Japanese social relationship. Up to the beginning of World War II, the Japanese family had certain distinctive features as a social group. First, the individual's interests were subordinated to the welfare of the family as a whole. Secondly, the members of the family were organised in a hierarchy – head of the family, eldest son, second son and so on – and, thirdly, the relationship existing between these members was not one of domination and submission; the head of the family staked his honour on assuring the livelihood of its members, who willingly co-operated with him in his efforts to this end.

This relationship of 'familism', through the peculiar development of capitalism in Japan, is found not only in the family unit, but throughout Japanese society, in rural communities and elsewhere. The same sort of relationship exists in labour relations: first, the interests of the individual are subordinated to those of the enterprise. Secondly, a distinct hierarchy is formed by the creation of differences in status within the enterprise. In the third place, although the rights of management are guaranteed by this hierarchical structure, only very rarely must this authority be openly exercised, since, in return for the company's emphasis on the provision of security for the workers, they come to have a very strong affection for the company. 'Managerial familism' is, I feel, a suitable term for this relationship.

Familism as an 'ideology of management' has been emphasised far less, however, since the end of World War II, when it developed into what is merely a policy of concern for the security of the workers' livelihood. There are also very wide differences in the form taken by this relationship between large and small enterprises. In small firms there is little wage variation based on length of service and age. Up to the age of thirty or so, wages rise in much the same way as in the larger enterprises, but after that the curve rises less steeply and a fairly wide gap develops between wages in the two types of enterprise. In small companies retirement benefit is either very small or non-existent, while welfare benefits, where they exist at all, are usually confined to token money gifts of congratulation or commiseration, requiring little regular expenditure.

However, what the small enterprise is unable to provide in material terms is made up for by the strong personal ties, since the smaller the enterprise the higher the proportion of family members among its workers. These family members serve as the strategic base from which management can adjust human relationships within the enterprise, the other employees being embraced in the 'quasi-family' atmosphere. In many cases the proprietor of a small company will live on or very near the company's premises and young employees will live in their employer's house – a literal 'management family'.

In a large firm, however, the ties of personal connection are inevitably slender, being replaced by systematised security in the form of wages, welfare benefits and training schemes. Nevertheless, although the form taken may differ, the attitude of management is essentially the same.

This system of management does, of course, have its shortcomings. Personnel management becomes very inefficient; this leads to confusion in the operation of the enterprise and makes it very difficult to express any of the operations in terms of figures. It is almost impossible to predict results and formulate rational plans for the future – a vital task in any modern enterprise. It is true, of course, that this type of management is gradually changing, in response to the demands of technical progress, and efforts are being made, slowly and gradually, to put personnel administration on a more scientific and ordered basis. However, this process has got no further than a partial reform and systematisation of these tradi-

tional Japanese practices and it cannot yet be said that it has achieved a real correction of the shortcomings referred to above.

Nevertheless, in spite of these shortcomings, the rapid progress made by Japanese industry over the past few years is undeniable; this type of managerial practice must have had some useful function to perform for it ever to have come into existence. We find that it has, in fact, performed a very useful function by arousing in the worker a feeling of affection for the company that employs him and by involving his daily life very closely with the fortunes of that company.

It might seem that these results have been bought at a very high price, but when we consider the number of working days that might be lost and the even higher wages and better working conditions that might have to be conceded as a result of labour disputes, as well as the fall in productivity that could result from low morale, it can be seen that the price is still very low. This is shown very clearly by an international comparison of labour disputes. According to Ross and Hartmann, Japan has the second highest figure – after France – for the ratio of the number of workers participating in strike action to the total number of organised workers.[1] Japan's figure is about the same as that for America, three to four times as high as that for Britain, and ten times as high as that for West Germany. However, if we consider the average number of working days lost per striking worker, excluding 1952, the figure for Japan is from three to five days, about the same as for Britain, and half the figure for West Germany. In America the figure is ten days, two or three times that for Japan. Further, if we divide the total number of days lost through strike action by the total number of organised workers (not the total labour force), the resulting figure – the number of working days lost per union member – is less than one for every year except 1952. In other words, fewer working days were lost than there were union members.

We can summarise this situation thus: in other countries, unions are very loath to indulge in strike action, but once they do go on strike, they fight to the bitter end, so that strikes tend to be protracted. In Japan, on the other hand, strikes occur relatively often, but end quickly, involving no very deep conflict between labour and management.

I

IV

Let us now consider the reasons for the development of these attitudes on the part of management, which in their turn have given rise to the characteristic pattern of Japanese trade union behaviour. Personnel management as we see it today in Japan is of fairly recent origin; until about 1920 it was very rare for a company to be the direct employer of its own workers. Until the Russo-Japanese War of 1904–5, most recruitment and disposition of the work force was done by experienced workers – who would be called foremen today; they were also responsible for training, so that there was no reason for the personnel administration organisation of the company to develop. Wages were paid on an efficiency basis, either to each individual, or calculated for a small group of workers together. The wage system did not therefore provide any security of livelihood as it does today, nor were welfare services at all highly developed.

Under such conditions, labour was very mobile indeed. According to Shokkō Jijō, the excellent government survey published in 1905, 12·9 per cent of machine operators had less than six months' service with their firm, and 39·6 per cent between six months' and one year's service; so over 50 per cent had changed their place of employment within the space of a year.[2] However, from about the time of World War I, the situation began to change, with workers coming under the direct control of the company which employed them, and by about 1920 this pattern of very high labour mobility had almost disappeared. The reason for this change was not only that a high labour turnover involved companies in heavy expenditure on recruiting, but that, with the introduction of modern techniques, the migratory 'jack-of-all-trades' type of worker was needed less and less, his place being taken by large numbers of workers who had received a training in a fairly narrowly specialised field, and were therefore better fitted for the new situation with its emphasis on the the division of labour.

From about this period, therefore, the enterprise organisation that we see today began to be established, with its status differentiation between staff, first-class operators, second-class operators and so on. The longer the worker's service with the firm, the better were his working conditions, the stress being shifted from degree of experience – hitherto the most important factor – to

length of service, with a corresponding increase in the welfare services provided. Thus working conditions, which had been relatively uniform in the period of high labour mobility, began to show differences from firm to firm and the system of personnel management which we have referred to above came into being.

We are still left, however, with the problem of why, once this development had occurred, there was not a far more thorough bureaucratisation of the system of personnel management, and why employment relations in Japan never took the shape of a true buying and selling of labour power. The explanation lies in the peculiarity of the development of the Japanese labour movement. Of course, as so often happens with origins, it is difficult to say exactly when the Japanese labour movement started. We are told that in 1862 there was a strike in part of the Besshi Copper Mine, or that at the beginning of the Meiji era there were riots at various mines and movements for the amelioration of working conditions in the handicraft industries. But these were for the most part lacking in organisation and somewhat different in character from the labour movement as we know it today – a by-product of the Industrial Revolution, whose participants are true modern workers. Looked at in this way, the strikes at the Amamiya Spinning Company in Kōfu in 1886 and the Temma Spinning Company in Osaka in 1894 can perhaps be regarded as the first labour disputes to take place in firms with modern systems of management, so that we can put the start of the Japanese labour movement in the last years of the nineteenth century.

The war with China in 1894–5 gave a great impetus to the development of capitalism in Japan. The labour force increased rapidly, to 400,000 in 1897 and 600,000 in 1907, while the rise in prices caused by the war and the cuts in wages in the depression immediately afterwards both affected the living standards of the people very seriously. It was in these circumstances that Katayama Sen, Takano Fusatarō and others who had returned from America set up the Shokkō Giyūkai (Volunteer Workers' Society) in 1897. That year also saw the first public lecture meeting on the subject of labour, with enthusiastic speeches pointing out the need for the workers to organise. As a result, the Machine Operators' Union was formed in 1897, followed in the next year by the Railwaymen's Union and the Printers' Union. Labour disputes also increased, from 32 cases in 1897, involving 3510 workers, to 43 cases involving 6293 workers in 1898.

This was the first peak in the history of the early labour movement, but it was not to last long. In 1900 the Public Peace Police Act was passed, with the purpose of suppressing the labour movement; it provided for terms of from one to six months' imprisonment or fines, for anyone guilty of demanding negotiations with employers, or of inciting others to join a union or go on strike. The police were thus given an excuse to interfere directly in labour disputes and, as a result, there began a series of defections by union members; this was a severe blow to the union movement, which had not yet achieved sufficient strength to weather an attack of this sort. In the next year, 1901, the first workers' political party, the Shakai Minshutō (Social Democratic Party), was founded by Abe Isoo, Katayama Sen, Kōtoku Shūsui and others, only to be suppressed on the day of its formation, leaving the workers with no legal means of bettering their condition. Inevitably the labour movement continued to work illegally, but it was everywhere faced with ruthless oppression.

From that time until World War I, the labour movement was given no chance to breathe, though sporadic disputes did occur. The chance for the second peak in its history came with the foundation in 1912 of the Yūaikai (Friendly Love Society). This society pledged itself to 'fight to the death for socialism' and espoused a creed of harmony between labour and capital. Nevertheless, the outbreak of World War I aided even further the development of capitalism – the work force increasing to 1,700,000 – and the distress caused by inflation led to an explosive increase in the number of labour disputes. In 1919, which saw 63,000 workers taking part in 497 disputes, the Yūaikai changed its name to Dai Nippon Rōdō Sōdōmei Yūaikai (Greater Japan General Federation of Labour – Friendly Love Society), becoming noticeably more militant, and in 1920 sponsored Japan's first May Day celebrations. In this period, too, the need began to be felt for a guiding principle for trade unionism and, partly because of the influence of the Russian revolution, the association between the labour movement and socialism began. In 1922 the Japan Communist Party was formed, and by 1926 Japan possessed 488 unions, with a total membership of over 350,000.

This was the state of the labour movement when, in 1925, the Public Peace Preservation Law was enacted. It laid down penalties

of up to ten years' imprisonment for anyone organising or joining any group which advocated the reform of the structure of the state, or the abolition of the system of private property. The maximum penalty for the first of these offences was later increased to one of death. In 1931 the number of persons arrested and interrogated under the provisions of this law was 10,422 and in 1932 the figure reached 14,624, clear evidence of the extent of the oppression. Thus trade union activity in Japan was once more cut off abruptly and entered its Dark Age, which was to last until trade unions were given public recognition after the end of World War II.

Two facts emerge clearly from an examination of the history of the trade union movement in Japan. The first is that the establishment of Japan's characteristic type of personnel management coincided with the second peak in the movement's history, and that these management methods were developed by employers, largely to combat the rising tide of union activity. The second fact that we should note is that in pre-war Japan the union movement was stifled before it had developed into a movement worthy of the name. The normal course of development is for the working class to mature as capitalism develops, becoming a true proletariat and organising itself into a union movement. In Japan's case, however, the organisation of the workers lagged far behind the development of capital, so that, when union activities really began after World War II, employers' management policies and methods were already prepared; they were thus able to seize the initiative in the very areas where the unions should have been most active.

In other countries industrial or craft unions with a highly centralised organisation have been able to monopolise their own specific sector of the labour market, controlling it and turning the relationship of supply and demand to their own advantage in order to improve working conditions. In addition, because of their control over the labour market, the unions have assumed the function of labour exchanges and, where union membership is voluntary, this has been a strong factor in encouraging workers to join the unions. And since financial difficulties arising from illness or unemployment can turn a union member into a strike-breaker, the function of unions as mutual aid societies is also very important. They have therefore undertaken the operation of hospitals and sales organisations, as well as the payment of benefits to their members on the occasion of sickness, death or unemployment. These activities have served to make the unions more attractive, increasing their appeal to their members.

In Japan, however, by the time the unions came into existence, the characteristic employment relationship of the Japanese enterprise had already developed, leaving the unions no scope for controlling the labour market and acting as employment agencies. The employers had also set up their own facilities to provide for the welfare and security of their workers, leaving the unions no direct point of contact with the daily lives of their members. The basis for the development of industrial or craft unions in Japan was thus very weak, and the unions have ended up by being enclosed within each firm as enterprise unions. In such circumstances, the only basis for united action is on political issues; as a result, the Japanese unions have inevitably tended to give too much weight to political activity. A situation exists, therefore, where unity is possible on political issues, but unattainable on issues which are more properly the concern of the trade union movement. The unions are therefore forced into the situation we have already mentioned, where they can act only after the initiative has been taken by the management.

There are gradual changes in this pattern of labour–management relations. As social and economic conditions alter and technical innovation proceeds, further changes are inevitable in the future, particularly if the unions are able to strengthen their position. But the description we have given presents a fair picture of the peculiar character of labour relations in Japan at the present moment.

Notes

1. A. M. Ross and P. T. Hartman, *Changing Patterns of Industrial Conflict* (London, 1960).
2. *The Condition of the Workers*, vol. 2, p. 11.

MAKING-OUT: A COUNTER-SYSTEM OF WORKERS' CONTROL OF WORK SITUATION AND RELATIONSHIPS

D. Roy

As part of a broader examination and appraisal of the application of piecework incentive to the production line of an American factory, this paper essays the simple but largely neglected task of

exploring the network of intergroup relations in which the work activity of machine operatives is embedded. Exploration will be restricted to a limited sector of the total web of interaction in one shop; description will centre upon those relationships that provide support to the operator group in its resistance to and subversion of formally instituted managerial controls on production. It is hoped that observations reported here not only will bear upon the practical problem of industrial efficiency but will also contribute to the more general study of institutional dynamics.

This could be considered the third in a series of attempts to make more careful discriminations in an area of research that has been characteristically productive of sweeping generalisations, blanket conceptualisations, or algebraic gymnastics that tend to halt inquiry at the same time that they lay a fog over otherwise easily discerned reality. Data for all three papers were acquired in an investigation of a single work situation by a single technique of social inquiry, participant observation. The writer was employed for nearly a year as radial-drill operator in one of the machine shops of a steel-processing plant, and he kept a daily record of his observations and experiences relating to work activity and social interaction in the shop. His major interest lay in the phenomenon of restriction of output, or 'systematic soldiering', the practice of which various sociological soundings have revealed in the lower depths of our industrial organisation. To complete the analogy: the writer donned a diving suit and went down to see what it looked like on the bottom.

One conclusion has already been set forth, namely, that the usual view of output restriction is grossly undifferentiating.[1] Different kinds of 'institutionalised underworking' were practised, each with its characteristic pattern of antecedents and consequences. The blanket term 'restriction' was found to cloak all-important contrarieties of work behaviour. Machine operatives not only held back effort; sometimes they worked hard. The very common failure to note such contrarieties has tended, of course, to impede the progress of research by checking consideration of the specific conditions under which differences in behaviour occur.

A second finding was the discovery of complexity where simple lines of relationship had generally been assumed to exist.[2] When inconsistencies in the operator's behaviour seemed to contradict the hypothesis that variations in application of economic incentive could account for the variations in work effort, a more intensive examination or response to piecework was undertaken. This disclosed that piecework incentive was not equivalent to economic incentive and that attainment of piecework 'quotas' afforded machine operators a complex of rewards in which the strictly economic might or might not play a part.

The third set of observations, to be here discussed, again exhibits complication in a picture that has come to be accepted as simple in design. Here the focus of interest is the structure of 'informal' intergroup connections that bear directly upon work behaviour at the machine level. The material will not deny the hypothesis that the willingness of operatives to put forth effort is a function of their relationship with management or the widely held affirmation that this relationship is mediated by the organisation of operatives into 'informal groups'. It will indicate, however, that further advances in the understanding of work behaviour in the factory may involve attention to minor as well as major axes of intergroup relations. It will show that the relevant constituents of problematic production situations may include 'lateral' lines of interaction between sub-groups of the work force as well as 'vertical' connections between managerial and worker groups.

It will be seen, in other words, that the interaction of two groups in an industrial organisation takes place within and is conditioned by a larger intergroup network of reciprocal influences. Whyte has called attention to the limitations of studying groups in 'isolation', without regard for the 'perspectives of large institutional structures'.[3] A second warning might be: The larger institutional structures form networks of interacting groups.

As a bona fide member of an informal group of machine operatives, the writer had an opportunity to observe and experience management–work group conflict in its day-to-day and blow-by-blow particulars. Also, he participated in another kind of social process, intergroup co-operation. Not only did workers on the 'drill line' co-operate with each other as fellow-members of a combat team at war with management; they also received considerable aid and abetment from other groups of the shop. This intergroup co-operation was particularly evident when operators were trying to 'make out', or attain 'quota' production, on piecework jobs.

It has been noted in another connection that machine operators

characteristically evinced no reluctance to put forth effort when they felt that their group-defined piecework quotas were attainable.[2] It might seem, at first glance, that the supporting of operators during intensive application to 'getting the work out' would represent co-operation *with* and not *against* management. However, the truth is that operators and their 'allies' joined forces in certain situations in a manner not only unmistakably at variance with the carefully prepared designs of staff experts, but even in flagrant violation of strongly held managerial 'moral principles' of shop behaviour. In short, machine operators resorted to 'cheating' to attain their quotas; and since this often involved the collusion of other shop groups, not as mere 'accessories after the fact' but as deeply entangled accomplices, any managerial suspicion that swindling and conniving, as well as loafing, were going on all the time was well founded. If the workers' conviction that the echelons of management were packed with men addicted to the 'dirty deal' be additionally considered, it might appear that the shop was fairly overrun with crooks. Since a discussion of 'contrast conceptions'[4] cannot find a place within the limited scope of this paper, it must suffice at this point merely to declare that the kind of effort made by operators and their aids to expedite production, when they did try to expedite it, was actually in many respects in conflict with management.

One belief, universally accepted in the work group, may be phrased thus: 'You can't "make out" if you do things the way management wants them done.' This gem of shop wisdom thus negatively put is hardly a prescription for action, but its obverse, 'You've got to figure the angles', gave all hands plenty to do.

According to Al McCann (all names used are fictitious), the 'Fagin' of the drill line, 'They time jobs to give you just base rates. It's up to you to figure out how to fool them so you can make out. You can't make any money if you run the job the way it's timed.'

We machine operators did 'figure the angles'; we developed an impressive repertoire of angles to play and devoted ourselves to crossing the expectations of formal organisation with perseverance, artistry and organising ability of our own. For instance, job timing was a 'battle all the way' between operators and time-study men. The objective of the operators was good piecework prices, and that end justified any old means that would work. One cardinal principle of operator job-timing was that cutting tools be run at lower speeds and 'feeds' than the maximums possible on subsequent production, and there were various ways of encouraging the institution of adequate differentials. Also, operators deemed it essential to embellish the timing performance with movements only apparently functional in relation to the production of goods: little reachings, liftings, adjustings, dustings and other special attentions to conscientious machine operation and good housekeeping that could be dropped instanter with the departure of the time-study man.

However, the sophistication of the time-study men usually matched the strategy employed against them. The canniest operators often gave of their best in timing duels only to get 'hopeless prices' for their pains:

> Gus Schmidt was timed early in the evening on a job, and given a price of $1.00 per 100 for reaming one hole, chamfering both sides of three holes, and filing burrs on one end of one hole. All that for one cent!
>
> 'To hell with them,' said Gus.

This is not to say that the 'hopeless price' was always truly hopeless. Since the maintenance of an effective control over job-timing and hence price-setting was an uncertain, often disheartening matter, operators were forced to develop skills for turning bad into good. Under the shaping hands of the 'angle-applicators' surprising metamorphoses sometimes took place. Like the proverbial ugly duckling that finally feathered out into a beautiful swan, piecework jobs originally classified in operator vernacular as 'stinkers' came to give off the delightful aroma of 'gravy'. Without going into the particulars of the various types of operation, one might say that jobs were 'streamlined'. This streamlining was, of course, at times 'rough on the tools' and adverse in its effects on the quality of output. The jettisoning of quality called, necessarily, for a corresponding attention to ways and means of shielding supervisors and inspectors from discovering the sacrifices and consequently brought into further play the social graces of equivocation, subterfuge and prestidigitation.

Still, the adroitness of the machine operators, inventing, scheming and conniving unto themselves to make quotas attainable, was not enough. Many 'stinkers' would not yield before the whitest heat of intelligence or the most cavalier disregard for

company property. An appreciable incidence of failure should not be surprising when it is kept in mind that the black arts of 'making out' were not only responses to challenge from management but also stimulations, in circular interaction, to the development of more effective countermagic in the timing process. It would be hard to overestimate the wizardry of the time-study men with pencil and paper in computing 'angle-tight' piecework prices. During the latter months of his employment, months that marked the peak of his machine performance, the writer was able to achieve quota earnings approximately half the time that piecework jobs were offered. If this experience is roughly representative of the fortunes of the drill-line group, the battle with the stopwatch men was nip and tuck.

It is to be expected that a group of resourceful operatives, working with persistent intent to 'make out' at quota levels, and relying heavily upon illegal practices, would be alert to possibilities of assistance from groups that were able and willing to give it and would not hesitate at further flouting the rules and regulations in cultivating it. It is also to be expected that the upholders of a managerial rational and moral order would attempt to prevent corruptive connections and would take action to stamp out whatever subversive organisation did develop. During the eleven-month study, machine operators, including the drill-line men, were enjoying the co-operation of several other shop groups in an illegal facilitation of the 'make-out' process. This intergroup network effectively modified certain formally established shop routines, a too close attachment to which would handicap the operators. The 'syndicate' also proved adequate in circumventing each of a series of 'new rules' and 'new systems' introduced by management to expurgate all modifications and improvisations and force a strict adherence to the rules.

The shop groups that conspired with the operators were, namely, the inspectors, the tool-crib men, the time-checkers, the stock men and the set-up men. With a single exception, these 'service' groups stemmed from lines of authority distinct from the one for which the operators formed the base. The one exception was the set-up group; it was subordinate to the same set of officials in the 'production' line of authority that controlled the operators. A brief description of the duties of each of these service groups and a rough tracing of the sequences of interaction involved in the prescribed work routine of the drill men will indicate the formal pattern of intergroup relations within which informally instituted variations were woven.

The set-up men

A chief function of the set-up men was to assist machine operators in the 'setting-up' of jigs and fixtures preparatory to operation of machines in the processing of materials. It included the giving of preliminary aid and advice at the beginning of the production process, at which time the set-up men would customarily 'run the first piece' to show operators how to do it and to indicate that the set-up was adequate to meet work specifications. The duties of the set-up men also included 'trouble-shooting' on occasions when operators encountered difficulties that effected a lowering of the quality of output below inspection standards or a reduction of the rate of output unsatisfactory to operators or supervisors.

The inspectors

The chief function of the inspectors was to pass judgement on the quality of the output of the machine operators, either accepting or rejecting work turned out, according to blueprint specifications. Their appraisals came at the beginning of operations, when especially thorough examinations of the first pieces processed were made, and subsequently at varying intervals during the course of a job.

The tool-crib men

The tool-crib attendants served the operators as dispensers of jigs, fixtures, cutting tools, blueprints, gauges and miscellaneous items of equipment needed to supplement basic machinery and operator-owned hand tools in the processing of materials. They worked inside a special enclosure conveniently located along one of the main arterials of shop traffic and did most of their dispensing across the wide sill of a 'window', an aperture which served, incidentally, as locus of various and sundry transactions and communications not immediately relevant to tool-dispensing. There were two other openings into the crib, a door, two steps from the window, and a wide gate, farther down the corridor.

The stockmen

The stockmen were responsible for conducting a steady flow of materials to the machines for processing. Their work called for the

removal of finished work as well as the moving-up of fresh stock and involved a division of labour into two specialisations: 'stock-chasing' and 'trucking'. The chief duties of the stock-chasers were to 'locate' unprocessed materials in the various storage areas, when operators called for stock, and to direct the activities of the truckers, who attended to the physical transportation.

The time-checkers

The time-checkers worked in another special enclosure, a small 'time cage', from which they distributed to the operators the work orders 'lined up' by the schedulemen of the Planning Department and within which they registered the starting and completion times of each job. There were four time-registering operations for every work order. First, upon presenting an operator with a work-order slip, the checker would 'punch' him 'on set-up' by stamping a separate order card with a clocking mechanism that registered the hours in tenths. Later, when the operator would call at the cage window to announce completion of all preparatory arrangement for the actual processing of materials, the checker would punch him 'off set-up' and 'on production'. Finally, following another operator announcement, the checker would clock the termination of the machining process with a fourth punch. At the time of his terminal punch the operator would report the number of 'pieces' completed on the job just concluded and would receive a new work order to start the cycle over again. And since the terminal punch on the completed job would be registered at the same time as the initial punch on the new one, hours on shift would be completely accounted for.

OPERATOR INTERACTION WITH SERVICE GROUPS

The machine operator's performance of each individual job or order assigned to him involved formal relationships with service groups in well-defined sequences or routines.

First, the operator would receive his work order from the time-checker. Next, he would present the work order to a tool-crib attendant at the crib window as a requisite to receiving blueprints, jigs, cutting tools and gauges. At the same time, that is, immediately before or after approaching the crib attendant, sometimes while waiting for crib service, the operator would show his work order to a stock-chaser as a requisite to receiving materials to work on. The stock-chaser, after perusing the order slip, occasionally with additional reference to the blueprint, would hail a trucker to bring the necessary stock to the operator's machine. If there were no delay in contacting a stock-chaser or in locating and moving up the stock, a load of materials would await the operator upon his arrival at his machine with equipment from the tool crib.

Upon returning to his machine, the operator would proceed with the work of 'setting up' the job, usually with the assistance of a set-up man, who would stay with him until a piece was turned out whose quality of workmanship would satisfy an inspector. In appraising a finished piece, the inspector would consult the blueprint brought from the crib for work specifications and then perform operations of measurement with rules, gauges, micrometers or more elaborate equipment. The inspector might or might not 'accept' the first piece presented for his judgement. At any rate, his approval was requisite to the next step in the operator's formal interactional routine, namely, contacting the time-checker to punch 'off set-up' and 'on production'.

The operator would ordinarily have further 'business' contact with a set-up man during the course of production. Even if the job did not 'go sour' and require the services of a 'trouble-shooter', the set-up man would drop around of his own accord to see how the work was progressing. Likewise, the operator would have further formal contact during the course of his job with inspectors and tool-crib attendants. Each inspector would make periodic 'quality checks' at the machines on his 'line'; and the operator might have to make trips to the tool crib to get tools ground or to pick up additional tools or gauges. He might also have to contact a stock-chaser or truckers for additional materials.

Upon completion of the last piece of his order the operator would tear down his set-up, return his tools to the tool crib and make a final report to the time-checker. Should the job be uncompleted at the close of a shift, the operator would merely report the number of pieces finished to a checker, and the latter would register a final punch-out. The set-up would be left intact for the use of the operator coming in to work the next shift.

MAJOR JOB CATEGORIES

Certain variations in types of jobs assigned to operators are pertinent to a discussion of intergroup collusion to modify formal work routines. These variations could be classified into four categories: (1) piecework; (2) time study; (3) rework; and (4) set-up.

Each piecework job carried a price per 100 pieces, determined by the timing operations mentioned earlier. Time-study and rework jobs carried no prices. The time-study category included (*a*) new jobs that had not yet been timed and (*b*) jobs that had once carried a piecework price. As the label indicates, rework jobs involved the refinishing of pieces rejected either by inspectors or in the assembly process but considered salvageable by reprocessing.

Since time-study and rework jobs carried no piecework prices, operators engaged in these two types of work were paid 'day rates', that is according to an hourly base rate determined in collective bargaining. The base rates represented minimal wage guarantees that not only applied to 'day work' but also covered piecework as well. If an operator on piecework failed to exceed his base rate in average hourly earnings on a particular day, he would be paid his base rate. Failure to produce at base rate or above on the first day of a piecework job did not penalise an operator in his efforts to earn premium pay on the second day; nor did failure to attain base rate on one piecework job on a given day reduce premiums earned on a second job performed that day.

Not a fourth type of job, but measured separately in time and payment units, were the set-up operations. Piecework jobs always carried piecework set-ups; failure to equal or exceed base rate on set-up did not jeopardise chances to earn premium on 'production', and vice versa. Time-study jobs frequently carried piecework set-ups; rework never.

Obviously, these formal work routines may easily be modified to fit the perceived needs of machine operators. Possibilities for the development of 'make-out angles' should be immediately apparent in a work situation characterised by job repertoires that included piecework and day-work operations; minimum-wage guarantees uniform for all work done; and separate payment computations by jobs and days worked. If, for instance, time formally clocked as day work could be used to gain a 'head start' on subsequent piecework operations, such a transferral might mean the difference between earning and not earning premiums on doubtful piecework jobs. Similarly, time on 'hopeless' piecework jobs might be applied to more promising operations; and the otherwise 'free time' gained on 'gravy' jobs might be consumed in productive anticipation of the formal receipt of ordinarily unrewarding piecework. Especially lush 'gravy' jobs might even contribute extra time enough to convert 'stinkers' into temporary 'money-makers'. Realisation of such possibilities in any given case would necessarily involve obtaining, without a work order, the following: (1) identification of future operations as listed in sequence on the schedule board inside the time cage; (2) jigs, blueprints and cutting tools appropriate to the work contemplated; (3) stock to work on; (4) set-up help and advice; (5) inspection service; and (6) 'trouble-shooting' assistance as needed. Obviously, this sequence of accomplishments would call for the support of one or more service groups at each step. That the required assistance was actually provided with such regularity that it came to be taken for granted, the writer discovered by observation and personal experience.

The following diary recording of interaction between the writer and a time-checker may be indicative of the extent to which service-group collaboration with the operators in perverting the formal system of work routine had become systematised:

> When I came to punch off the rework, the time-cage girl said, 'You don't want to punch off rework yet, do you?' – suggesting that I should get a start on the next job before punching off rework.

Even line foremen, who in regard to intergroup collusion preferred the role of silent 'accessory after the fact', became upset to the point of actual attempted interference with formal rules and regulations when the naïve neophyte failed to meet the expectations of his own informal system.

> Art [foreman] was at the time cage when I punched off the day work of rereaming and on to the piecework of drilling. He came around to my machine shortly after.
>
> 'Say,' he said, 'when you punch off day work on to piecework, you ought to have your piecework already started. Run a few;

then punch off the day work, and you'll have a good start. You've got to chisel a little around here to make money.'

Acceptance of such subversive practices did not extend, however, to groups in management other than local shop supervision. The writer was solemnly and repeatedly warned that time-study men, the true hatchet men of upper management, were disposed to bring chisellers to speedy justice.

Gus went on to say that a girl hand-mill operator had been fired a year ago when a time-study man caught her running one job while being punched in on another. The time-study man came over to the girl's machine to time a job, to find the job completed and the girl running another.

NEW RULES AND NEW SYSTEMS

During the near-year that he spent in the shop the writer felt the impact of several attempts to stamp out intergroup irregularities and enforce conformity to managerial designs of procedure. He coincidentally participated in an upholding of the maxim: 'Plus ça change, plus c'est la même chose.'

Attempts to tighten controls came in a series of 'new rules' or 'new systems' promulgated by bulletin-board edicts. How far the beginning of the series antedated the writer's arrival is not known. Old-timers spoke of a 'Golden Age' enjoyed before the installation of the 'Booth System' of production control; then operators 'kept their own time', turning in their work orders as they saw fit and building 'kitties' on good jobs to tide them over rainy days on poor jobs.

The first new rule during this study went into 'effect' less than two months after the writer was hired. It was designed to tighten controls in the tool-crib sector, where attendants had not only been passing out set-ups ahead of time but allowing operators or their set-up men to enter the toolroom to make the advance pick-ups themselves. An aim of the new rule was also to curb the operators' practice of keeping 'main set-ups' at the machines instead of turning them in at the completion of operations.

A new crib ruling went into effect today. A memorandum by Bricker [superintendent] was posted on the side of the crib window. Those who check out tools and jigs must sign a slip in triplicate, keeping the pink one and turning it in with the tools in exchange for the white original, which would constitute proof that the tools had been returned. No new set-ups would be issued until the old ones had been turned in.

An optimistic perception of the new procedures was expressed by young Jonesy, a tool-crib attendant and otherwise willing conniver with the operators: 'Tools are scattered all over the shop. This way we'll have them all in order in the crib, and the fellows can get them any time they need them.'

But multiple-drill operator Hanks, old-timer on the line, drew upon his lengthy experience with managerial efficiency measures and saw the situation differently:

Hanks commented unfavourably on the new ruling. He and the day man [his machine partner on the other shift] had been keeping the tools for their main set-ups at their bench, or, rather, under it. This practice, according to Hanks, was to ensure their setting up promptly without inordinate waste of time and to ensure their having all the tools needed. Hanks said that on a previous occasion he was told to turn in one of his main set-ups, which included over a dozen drills, reamers, taps, etc., of varying sizes. He did so, but, when he needed this set-up again, the crib man couldn't locate all the tools. He asked Hanks to come back in the crib and help him find them. Hanks refused. After several hours of futile search, Hanks was finally induced to 'come back and find his tools'. He did so on condition that it would not be on his own time. The foreman agreed to this.

'The same thing is going to happen again,' predicted Hanks. 'And I'm not going back there to find my tools they scatter all over, on my own time.'

Though the operators went through the formality of an exchange of slips when they exchanged set-ups, the new procedures did not modify the practice of getting set-ups from the crib ahead of time. Appreciable effects of the new ruling included making more paperwork for crib attendants at the same time that more work at assembling set-ups was thrust upon them. Jonesy's happy prediction did not materialise: the tools were not 'always in order'. Subsequent events confirmed Hanks's gloomy forebodings:

It took Paul [set-up man] and me several hours to get set up for the sockets, as the set-up given was incomplete.

Some time was spent in looking for an angle plate that was specially made for the job. Both Paul and Steve [superintendent] were irritated because the crib men could not find the plate.

We spent an hour setting up because we could not find the jig.

Included in the new ruling was a stipulation that blueprints and gauges be turned in by the operators at the end of each shift, though set-up paraphernalia other than prints and gauges were to be left at the machines as long as jobs were in operation. Calling for prints and gauges at the beginning of the shift, waiting at the crib window in the line that naturally formed, even when these items were 'located' immediately, consumed operator time.

Owing to the new crib ruling, he [Joe Mucha, the writer's machine partner on another shift] turned in the tap gauge. I spent twenty minutes trying to get it back again. The crib man could not find it and claimed that Joe had not turned it in. Joe stayed after 3 o'clock to help me get it, countering the arguments of the crib with the slip that he retained as evidence. Finally the gauge was located in the crib.

I started out a half-hour late on operation fifty-five on the pedestals, due to delay at the crib waiting to check out the print and gauge that Joe had just turned in.

Four months later the new crib ruling was modified by another that cancelled the stipulation regarding the turning-in of blueprints and gauges and called for changes in the paperwork of operator–crib-attendant relations. These changes were featured by a new kind of work order, duplicates of which became involved in tool-crib book-keeping. The change reduced the waste of operator time at the start of shifts, but to the burden of the crib attendants paperwork irritations were now added.

When I punched in on the rework and asked Walt [crib attendant] for a print, he fumed a bit as he sought a duplicate of my new-type yellow work order in a new file of his.

'I haven't been able to find more than one in five duplicates so far,' he said. 'And there's supposed to be a duplicate for every one.'

Walt said tonight, when I presented him with a work-order card for tools, 'That makes the twelfth card I've had and no duplicate!'

The tool crib under the new system is supposed to have duplicate work orders in their file of all jobs given operators. These duplicates are to be put in the toolroom files as soon as they are put on the board; and the operators are to sign these duplicates when checking out set-ups.

The 'new system' did operate to handicap operators in that they were not to receive new set-ups from the crib until they received the new yellow work orders from the time cage to check with the duplicates in the crib. However, set-up men roamed at will in the toolroom, grinding tools and fixing jigs, and were able to help the operators by picking up set-ups ahead of time for them. Their detailed knowledge of the various set-ups made it possible for them to assemble the necessary tools without the use of set-up cards.

'This is a good job,' I said to McCann [now set-up man]. 'I wish I could get it set up ahead of time, but I guess it's no use trying. I can't get the set-up now from the toolroom until I get the new work order from the time girls.'

McCann thought a moment. 'Maybe I can get the jig and tools out of the crib for you.'

McCann did get the jig and tools, and I got a half-hour's head start on the job.

The writer had found Ted, a stock-chaser, and his truckers, George and Louie, willing connivers in the time-chiselling process. They moved up stock ahead of time, even after the new system made presentation of the new work order to the stock-chaser a prerequisite to getting stock. Contrary to first impressions, for all practical purposes the situation was unchanged under the new system.

I could not go ahead with the next order, also a load of connecting rods, because the new ruling makes presentation of a work order to the stock-chaser necessary before materials can be moved up. So I was stymied and could do nothing the rest of the day.

About an hour before I was to punch off the connecting rods, I advised Ted that I would soon be needing another job. He immediately brought over a head of reservoir casings.

The new system also included complication of operator–inspector relations. Inspectors were now to 'sign off' operation from completed jobs before new work orders could be issued at the time booth. The 'signing-off' process included notation by the inspector of the time of operation completion, a double check on the time-checker's 'punch-out'. This added, of course, to the paperwork of inspectors.

Drill-man Hanks's first response to this feature of the new system was 'individualistic':

> Hanks commented on the new system tonight. He thinks that its chief purpose is to keep the operators from getting ahead on an operation and starting the next job on saved time. He said that the inspector checked him off a job tonight at 4.40, and he was not due to punch in on the next one until 6.10. He changed the time recorded by the inspector on his work slip to 6.10 and went ahead as usual. If he had not done so, there would have been a 'gap' of an hour and a half unaccounted for in the records.

The writer found himself 'stymied' at first but soon discovered that the new obstacle could be overcome without engaging in such a hazardous practice as 'forging'.

> It was 10 o'clock when we were ready to punch off set-up, and Johnny (set-up man) asked Sam (inspector) to sign me off set-up earlier, so that I could make out on set-up.
>
> 'Punch me off at 9 o'clock,' I said, not expecting Sam to check me off earlier, and purposely exaggerating Johnny's request.
>
> Sam refused. 'I can't do that! If I do that for you, I'll have to do it for everybody!'
>
> Sam seemed somewhat agitated in making the refusal.
>
> A few minutes later he said to Johnny, 'Why did you ask me to do that when Hanks was standing there?'
>
> Hanks had been standing by my machine, watching us set up.
>
> 'I can't take you off an hour back. Go find out when you punched in on this job in the first place.'
>
> Johnny consulted the time-cage girl as to the time I punched in on the job, later talked to Sam at Sam's bench while I was working and came to me with the announcement that it was 'fixed up' so that I made out on set-up and was credited with starting production at 9.30. This gave me an hour and a half of 'gravy'.

By the time the 'new system' was a month old, Sam was not only doing this for everybody but actually taking the initiative:

> When I punched off set-up for the eight pieces, Sam asked me if I wanted him to take me off set-up at an earlier time in order that I might make out on the set-up. I refused this offer, as it wasn't worth the trouble for me to stop to figure out the time.
>
> Instead of looking at the clock when an operator asks to be taken off set-up, Sam usually asks the operator, 'When do you want to be taken off?'

No sooner had the shop employees adjusted to this 'new system' and settled down to normal informal routine than they were shocked by a new pronunciamento that barred admittance to the toolroom to all save superintendents and toolroom employees:

> A new crib ruling struck without warning today. Typewritten bulletins signed by Faulkner (shop manager) were posted on the toolroom door, barring admittance to all save the toolroom employees and the two departmental foremen (superintendents), Bricker and Steve. Other foremen and set-up men are not to be admitted without permission from Milton, toolroom supervisor.
>
> Hanks predicts that the new ruling won't last out the week.

Stimulated by Hanks's prediction, the writer kept an eye on the toolroom door. The rule seemed to be enforced.

> On one occasion tonight Paul [set-up man] asked Jonesy to let him into the crib; he was in a hurry about something. But Jonesy shook his head, and Paul had to wait at the crib window with the rest of us.

Johnny, the set-up man, predicted that the new ruling would be 'tough on' the tool-crib employees, not on set-up men.

> Johnny says that the new rule is going to be tough on grinders and crib attendants, because set-up men and foremen have been doing rework of the grinding and have made it easier for them by coming in to help themselves to tools, jigs, etc.
>
> Johnny says that the new rule suits him fine. Now he can just

stand at the window and talk, and let the toolroom employees do the work.

The line foremen seemed to take offence at the new 'exclusion act' and threatened reprisals to the crib attendants.

At quitting time I noticed Gil (line foreman) talking to Walt at the crib window. Gil seemed very serious; Walt was waving his arms and otherwise gesturing in a manner indicating rejection of responsibility. I didn't catch any words but gathered that Gill was voicing disapproval or warning, and after Gil left I said to Walt, 'Looks like you're behind the eight-ball now!'

I noticed that Walt's hair was mussed, and he looked a little wild. He denied that he was in any trouble whatsoever; nor was he worried about anything whatsoever.

'I'm just working here!' he exclaimed. 'I just go by the cards and beyond that, I've got no responsibility!'

I was curious as to what Gil had told him and asked Johnny later, on the way home. I had noticed that Johnny was standing near by when Gil was talking to Walt. Johnny said that Gil was telling Walt that from now on the crib was going to be charged with every minute of tool delay to the operators – that, if there was any waiting for tools, Gil was going to make out allowance cards charging these delays to the crib.

Contrary to Hanks's prediction, the new rule did 'last out the week', and crowds milled around the crib window.

The boys seemed very much disgusted with the slow service at the tool crib. They crowd around the window (always a crowd there) and either growl or wisecrack about the service.

It was at this time that Jonesy, erstwhile optimist and regarded by shop employees as the most efficient of the crib attendants, decided that he had 'had enough'. He transferred to the quiet backroom retreat of tool-grinding. But several days later, just ten days since the new rule was promulgated, the sun began to break through the dark clouds of managerial efficiency. Hanks's prediction was off by four days.

While I was waiting for tools at the crib window tonight, I noticed the jockey (turret-lathe man) dash into the tool crib through a door that was left ajar; he was followed soon after by Gil. Later, when the door was closed, Paul shook it and shouted to the attendant, 'Let me in!' He was admitted.

Steve [superintendent] called out, 'Hey!' when he saw the jockey go into the crib. When the jockey came out, he spoke to him, and the jockey joshed him back. Steve did not seem to be particularly put out about it.

Soon the boys were going in and out of the crib again, almost at will, and set-up men were getting set-ups ahead of time for operators, ignored by the crib attendants.

I noticed that Johnny and others seemed to be going in and out of the crib again, almost at will.

I noticed tonight that Johnny got into the tool crib by appearing at the door and saying to the attendant, 'Let me in!'

So much for Faulkner's order – until he makes a new one!

When I asked Walt for some jaws to fit the chuck I had found, he said, 'We've got lots of jaws back there, but I wouldn't know what to look for. You'd better get the set-up man to come back here and find you some.'

Walt said to me, 'I break the rules here, but not too much, just within reason to keep the boys on production.'

Faulkner's order still hangs at eye level on the crib door.

'So much for Faulkner's order!' The 'fix' was 'on' again, and operators and their service-group allies conducted business as usual for the remaining weeks of the writer's employment.

CONCLUSIONS

This rough sketch of the operation of one shop 'syndicate' has been no more than indicative of the existence of intergroup co-operation in the lower reaches of factory social structure. No attempt has been made here to account for the aid extended by service groups, though suggestion that this assistance might be part of a larger system of reciprocal obligations has been implicit. It is apparent, for instance, that tool-crib attendants benefited from their practice of admitting operators and set-up men to the toolroom to seek and pick up equipment.

A more complete picture of intergroup relations would include conflict, as well as co-operation, between operators and the various

service groups. It could be shown, if space permitted, that changes in relationship accompanied, in cyclical fashion, changes in basic conditions of work.

Furthermore, attention has not been drawn to intragroup role and personality variations in intergroup relations. Such additional discriminations and the questions that they might raise in regard to the study of institutional dynamics must be left for future discussion.

As for their possible bearing on practical industrial administration, materials presented here seem to challenge the view held in some research circles that the 'human' problem of industrial efficiency lies in faulty communication between an economically 'rational' or 'logical' management and 'non-rational' or 'non-logical' work groups. While nothing has been offered to deny linkage between communication and efficiency, observations reported here suggest examination of the stereotypes of the two parties.[5] And questioning the fitness of the stereotypes may lead to a more fruitful conceptualisation of the process that is reputedly in need of attention: communication.

Do we see, in the situation studied, an economically 'rational' management and an economically 'non-rational' work group? Would not a reversal of the labels, if such labels be used, find justification? Does it not appear that operatives and their allies resisted managerial 'logics of efficiency' because application of those 'logics' tended to produce something considerably less than 'efficiency'? Did not worker groups connive to circumvent managerial ukase in order to 'get the work out'? Did not Walt, for instance, break the rules 'to keep the boys on production'? May not the common query of industrial workers, 'What in the hell are they trying to do up there?' be not merely reflective of faulty communication but also based on real managerial inadequacy, quite apart from a failure in 'explanation'? May it not be assumed that managerial inefficiency is and has been for some time a serious problem to those who labour?

If managerial directives are not the guides to efficient action that they are claimed to be, then, perhaps, 'logics of efficiency' would be better designated as 'sentiments of efficiency'. When failure to 'explain' is additionally considered, perhaps bulletin-board pronunciamentos might properly be classified with the various exorcisms, conjurations and miscellaneous esoteric monkey-business of our primitive contemporaries.

If we conceive of 'logical' behaviour not as self-contained ratiocinative exercises but as intellectual operations in continuous reciprocal interplay with concrete experience, machine operators and their service-group allies would appear the real holders of 'logics of efficiency'. Like big-city machine politicians, they develop plans for action that, under given conditions of situational pressures, 'work'.

But this rejection of commonly held stereotypes cannot lead to mere reversal of invidious distinctions; the situation is far too complex for that. The group life that the writer shared was by no means devoid of 'sentiments'. To the contrary, operator interaction was rich in shared feelings, attitudes and practices not only of doubtful bearing on getting the work out but often undeniably preventing production maximisation. Nor can it be maintained that management, in applying its 'sentiments of efficiency', was always ineffective. Perhaps solutions to the human problem of industrial efficiency would best be expedited by abandoning altogether the use of contrasted caricatures handed down to us from a pre-industrial social class structure. Instead of concerning ourselves with such blind-alley issues as who is 'rational' and who is not, we might recognise with John Dewey that both intellectual and emotional activity are essentials of goal-directed behaviour and that the development of effective communication focusing on production goals is a matter of instituting interactional processes that engender ideas, sentiments and plans for action held in common.

Notes

1. D. Roy, 'Quota Restriction and Gold-bricking in a Machine Shop', *Amer. J. Sociol.*, vol. 57 (1952), 427–42.
2. — 'Work Satisfaction and Social Reward in Quota Achievement', *Amer. Sociol. Rev.*, vol. 18 (1953) 507–14.
3. W. F. Whyte, 'Small Groups and Large Organisations' in J. R. Rohner and M. Sherif (eds), *Social Psychology at the Cross-Roads* (New York, 1951).
4. L. Copeland, 'The Negro as a Contrast Conception', in E. T. Thompson (ed.), *Race Relations and the Race Problem* (Duke U.P., 1939); S. K. Weinberg, 'Aspects of the Prison's Social Structure', *Amer. J. Sociol.*, vol. 47 (1942) 717–26.
5. W. F. Whyte, 'Semantics and Industrial Relations', *Human Relations*, vol. 8 (1949) 1–7.

However efficient is the allocation of resources in the production of individual commodities, the economy as a whole cannot be efficient if it leaves a large margin of resources unused. In the early 1930s, the number of British workers unemployed was well over two million; the attempt to explain this led Keynes to develop a theory which departed radically from the traditional (or 'classical') approach. Professor Harry Johnson summarises it in the first extract.

The concept of 'full employment', as Professor Joan Robinson shows in the second extract, has some important 'ideological' overtones.

MONETARY THEORY AND KEYNESIAN ECONOMICS

Harry Johnson

WHAT is the subject-matter of monetary theory? What is monetary theory about? This question brings us directly to the heart of the Keynesian revolution. Monetary theory before the Keynesian revolution was concerned primarily with the theory of the price level, the determination of the general level of prices. Associated with the price level was the question of economic fluctuations, which were connected with movements in the price level through the effects of rising prices in redistributing income from rentiers to entrepreneurs, increasing profit expectations and stimulating investments. Falling prices, on the other hand, redistributed income from entrepreneurs to rentiers, reduced profit expectations and depressed investment.

It is to be observed that classical monetary theory was not directly concerned with the level of employment, which entered only incidentally in connection with economic fluctuations. The classical theorists assumed a tendency to full employment of the economic system as a consequence of flexibility in wages and prices, and attributed unemployment to rigidity of wages and prices. We might note that, contrary to what some Keynesians have alleged, the classical economists did not assume automatic full employment. In their system full employment was a deduction from wage–price flexibility. On the other hand, they did not – as the Keynesians have shown – adequately investigate the effects of wage–price flexibility on employment and wrongly applied the analysis of a single commodity or factor to the economy as a whole.

Nor did classical theory investigate directly the theory of the rate of interest, which was held to be determined by real forces and not by monetary forces. The classical theorists did, however, introduce monetary forces to explain price trends and fluctuations, for example in the well-known form of the analysis of the effects of a divergence between the real and the market rate of interest.

To summarise, the classical theory was mainly concerned with the determination of the price level and with economic fluctuations, and was not directly concerned with the level of employment and the theory of interest. The essence of the Keynesian revolution was to shift the subject-matter of monetary theory, placing the emphasis on the level of employment as the central subject of monetary theory and posing the determination of the rate of interest as a specifically monetary problem. In the Keynesian theory, monetary forces influence the equilibrium rate of interest and do not merely, as in classical theory, exercise a transitional influence on market rates during periods of disequilibrium. On the other hand, in Keynesian theory the theory of the price level is reduced to a relatively minor role. Keynes took the level of money wages as given (through measuring all his variables in wage units) and assumed a closed economy. On this assumption there could be only three causes of a change in the price level: a change in the level of employment, a change in the wage unit and a change in technique of production, that is, a change in productivity. Of these three, Keynes was not concerned with changes in productivity, and the use of the wage unit left only variation in the level of employment as a cause of changes in the price level.

In addition to reducing the theory of the price level to a minor role, Keynes practically removed economic fluctuations from the picture. Although Keynesian concepts have since provided the means of clarifying considerably the mechanism of economic

fluctuations and cycles, the theory, as Keynes presented it, was not concerned with cycles: it was a static theory.

In summary, Keynes shifted the determination of the level of employment and the rate of interest from real theory into monetary theory, and reduced the former subject-matter of monetary theory – the price level and the economic fluctuations – to a minor position. In what follows I shall be concerned with various arguments centring on the question whether Keynes was right or not in doing this; with subsequent developments of Keynesian concepts as a result of theoretical discussion and practical experience; and with an evaluation of the present position of Keynesian theory in the light of experience over the past twelve years.

First, let me summarise briefly Keynes's theory. Its essence was an attack on the classical assumption of a tendency towards full employment and substitution for it of the possibility of underemployment equilibrium. The argument centred on the independence of decisions to save from decisions to invest. In a modern economy decisions to save are taken by a large number of individual households and firms, but decisions to invest are taken by a specialised group of entrepreneurs. There is, therefore, no guarantee that the desire to save will find expression in investment, or that an increased desire to save will be matched by increased investment. Instead, the increased desire to save may run to waste through unemployment, the increased savings of some being offset by reduction in the savings of others brought about by increased unemployment, total saving being made to correspond with the amount of investment that entrepreneurs want to do by a reduction of aggregate income.

Now according to classical theory, this could not happen: it would be prevented by a change in the rate of interest which would reduce saving and increase investment so as to make them balance at full employment. But, Keynes argued, the classical theory of interest involved indeterminacy, because it neglected the dependence of saving on income. Through variations in the level of income, savings and investment could be equated at any level of the rate of interest, and would not necessarily be equated at the full-employment level of income. (I shall return to this question of indeterminacy later on.)

So much for Keynes's attack on the classical theory; let us now turn to his own theory. I shall begin with a simple version of that theory which may be put as follows: output and employment are determined by the demand for output, which comes from two sources – consumption demand and investment demand. Investment decisions, which depend on entrepreneurial profit expectations and other factors, are taken for granted at this stage. Consumption demand depends largely on the level of income (other influences, including the rate of interest, are recognised but dismissed from the argument) and when income rises, consumption is certain to rise but not as much as income – in other words, as income rises, saving rises.

On this assumption the level of income is determined by the level of investment. Investment is taken as given by factors lying outside the theory; income must be such that saving equals investment and income is the sum of the consumption from this level of income and investment.

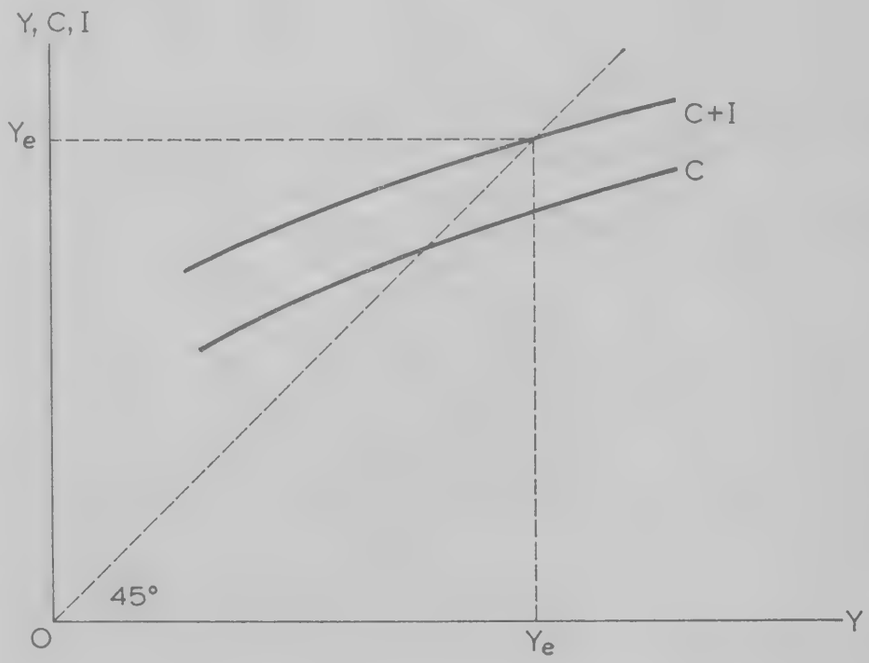


Fig. 14.1

This theory of income determination may be represented by two diagrams which are commonly employed in the literature: in the first (Fig. 14.1) we measure income and output (Y) along both axes and draw a 45° line; at any point on this line the quantities measured on the two axes are equal. We then draw a curve to represent the relation between consumption (C) and income. The vertical distance between this line and the 45° line represents saving (the unconsumed part of income). The diagram assumes that below some level of income the community will consume more than its income, that is, it will dissave; and that as income rises, consumption will increase but less than income. We now add the fixed amount of investment to the consumption that would take place at various levels of income, thus obtaining the curve $C + I$ which represents the demand for output at different levels of income. The intersection of this line with the 45° line determines the equilibrium level of income (Y_e). At a lower level of income and output demand for output will exceed output, and output and income will rise: conversely, at a higher level of income, demand for output will fall short of output, so output will fall. Another version of the same diagram is presented in Fig. 14.2, which plots saving (S) as a function of income; the equilibrium level of income is that for which saving is equal to the given level of investment (I). The second diagram is more commonly used and has been described by Schumpeter as the 'Keynesian cross' (by analogy with the Marshallian cross of the demand and supply curves).

This diagram shows the determination of the level of output and employment. For some purposes we are concerned with the relation between an initial change in demand for output and the total change in income to which it gives rise. This is the theory of the multiplier. Since an increase in investment will raise income and the increase in income will raise consumption, thereby raising the demand for output still further, the total increase in income will be some multiple of the initial increase in investment. In the simple theory just presented, the multiplier will be the reciprocal of the marginal propensity to save. In the early years after the publication of the *General Theory*, a great deal of the literature was devoted to elaboration of the theory of the multiplier. This literature, however, is not relevant to my main theme, so I shall not attempt to deal with it.

What I have just presented is a very simple version of the Keynesian theory. It is incomplete and unsatisfactory because it leaves both the rate of investment and the rate of interest undetermined. In the full version of the Keynesian theory, the rate of investment is determined by the balancing of the rate of return on investment, which Keynes termed 'the marginal efficiency of capital', against the rate of interest, the rate of investment being determined by the condition that, at the margin, the rate of interest and the marginal efficiency of capital should be equal. The rate of interest, in turn, is determined by the quantity of money and the demand for it, which Keynes analysed under the description of 'liquidity preference'.

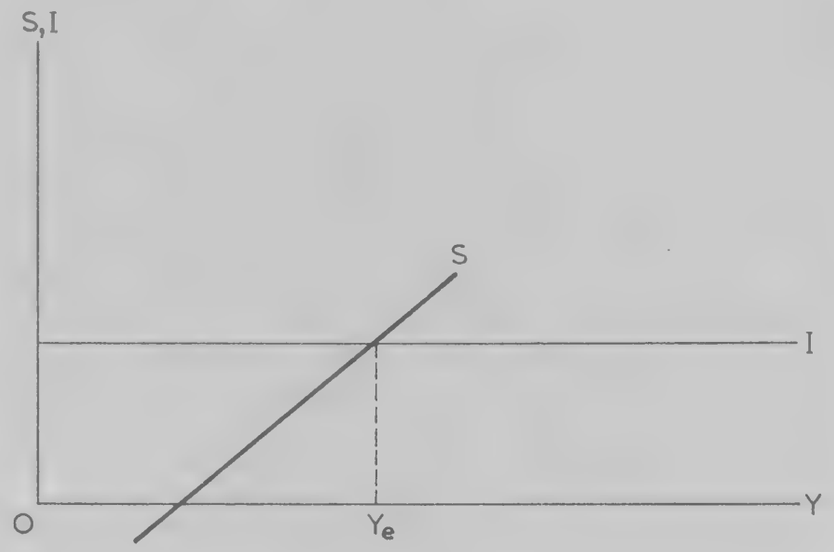


Fig. 14.2

The quantity of money is assumed to be determined by the monetary authorities. The demand for money is divided into two parts – the transactions demand (including the demand for precautionary balances) which depends on the level of money income, and the speculative demand. The speculative demand arises from the fact that, if the rate of interest is expected to rise, an owner of cash may do better not to invest it in securities im-

mediately, but to wait until he can buy those securities at a lower price (higher yield). The lower the actual rate of interest relative to some 'normal' level, the less the sacrifice of interest in not investing immediately, and the more likely a rise in the rate in the near future; consequently the greater the relative attraction of holding cash, and the greater the demand for money to hold as speculative balances. (Keynes, and even more so his followers, tended to write of the speculative demand for money as depending only on the actual rate of interest, and varying with it; this has given rise to confusion and misunderstanding, since it leaves implicit the role of expectations of future rates.) Keynes also held that at some low rate of interest the demand for idle balances might become completely interest-elastic, thus putting a bottom stop to the rate of interest.

Notice that this theory seems to imply a one-way chain of causation: the rate of interest is determined by the demand for and supply of money; the level of investment is determined by the rate of interest and the marginal efficiency of capital; and the level of income and consumption is determined by the level of investment and the propensity to save. There is, apparently, no reverse influence of saving and investment on the rate of interest; and this feature has led to much discussion and criticism of Keynes's theory. But the appearance is misleading: in fact, the Keynesian system is an interdependent system, in which all the variables influence each other, for the reason that the transactions part of the demand for money depends on the level of income, which in turn depends on the desires to save and invest. Appreciation of this interdependence dispels much of the startling novelty of Keynes's theory of interest as he presented it.

The full Keynesian theory, in which the interest rate, income, saving and investment, demand for money and supply of money are all mutually interdependent, can be represented very conveniently by a diagram developed by Professor J. R. Hicks,[1] which has proved so useful that it has become a standard tool of monetary theory.

Fig. 14.3A represents the relation between saving and income, and Fig. 14.3B the relation between investment and the rate of interest. From these two relations may be deduced the relation between the rate of interest and the level of income which will ensure equality between saving and investment, that is, will maintain equilibrium between the demand for and supply of output. This relation, known as the IS curve, is shown in Fig. 14.3C; it represents the requirements of equilibrium in the 'real' sector of the economy. The IS curve will be downward-sloping, since at a lower rate of interest investment will be greater, saving must be greater, and consequently income must be greater.

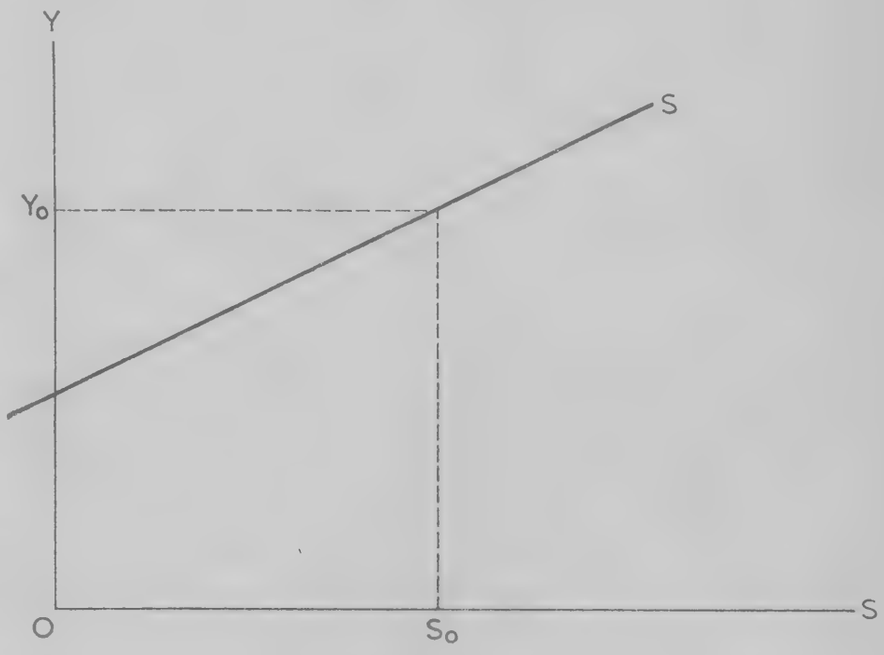


Fig. 14.3A

Fig. 14.4A represents the relation between the demand for and the supply of money. The upper right-hand quadrant represents the speculative demand for money, or the demand for speculative balances (M_2) as a function of the rate of interest. The lower left-hand quadrant represents the transactions demand for money (M_1), on the assumption that the transactions requirement is a fraction k of national income. The actual amount of money in existence $M(= M_1 + M_2)$ is represented by a bar, which can be shifted along the M_1M_2 axis to derive the relation between the rate of interest and the level of income at which the demand for money

will just equal the amount of it available, that is, which will maintain equilibrium between the demand for and supply of money. This relation, known as the LM curve, is shown in Fig. 14.4B; it represents the requirements of equilibrium in the 'monetary' sector of the economy. The LM curve will be upward-sloping, since at higher levels of national income more money will

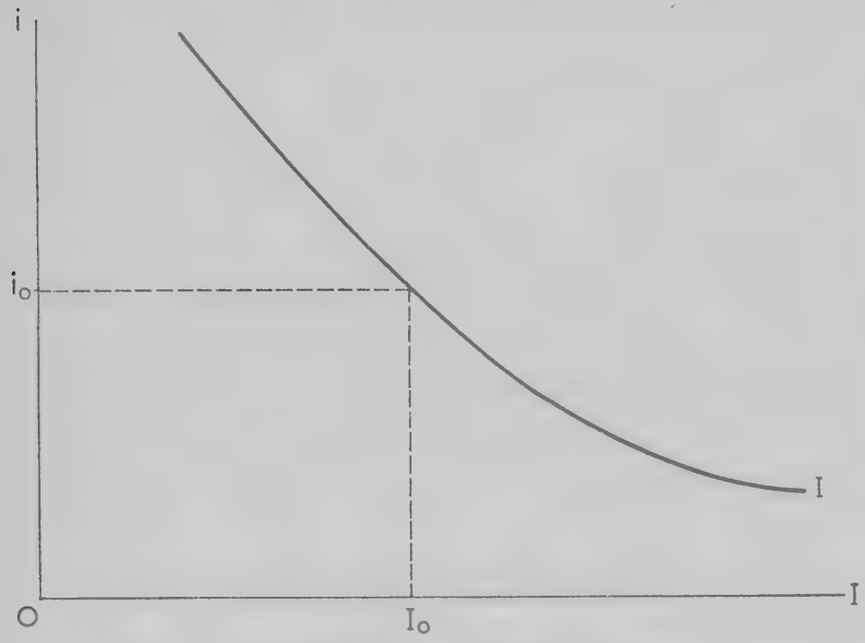


Fig. 14.3B

be required for transactions, leaving less available for idle balances, and to induce speculators to hold less idle money, the rate of interest must rise.

The IS and LM curves are combined in Fig. 14.5, which shows the determination of the rate of interest and the level of income at the levels which ensure simultaneous equilibrium in the monetary and real sectors. This diagram may be described as 'the Hicksian cross' after its inventor; it can be adapted to take account of more general assumptions than Keynes made – for example, that saving depends partly on interest rates and investment partly on income, and that both demands for money depend on both income and the interest rate; and it can be used to expound a number of important propositions and solve a number of problems in monetary theory.

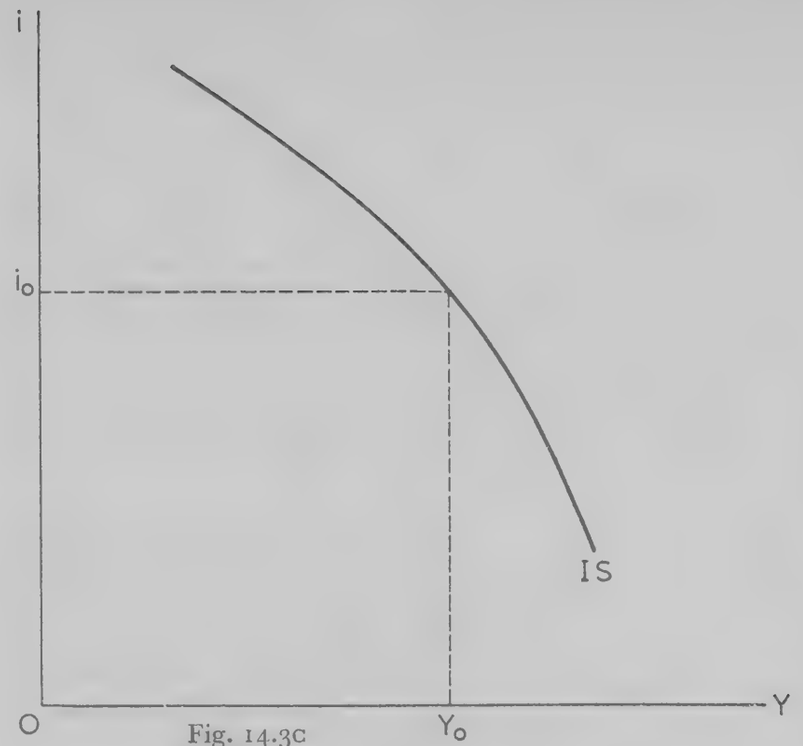


Fig. 14.3C

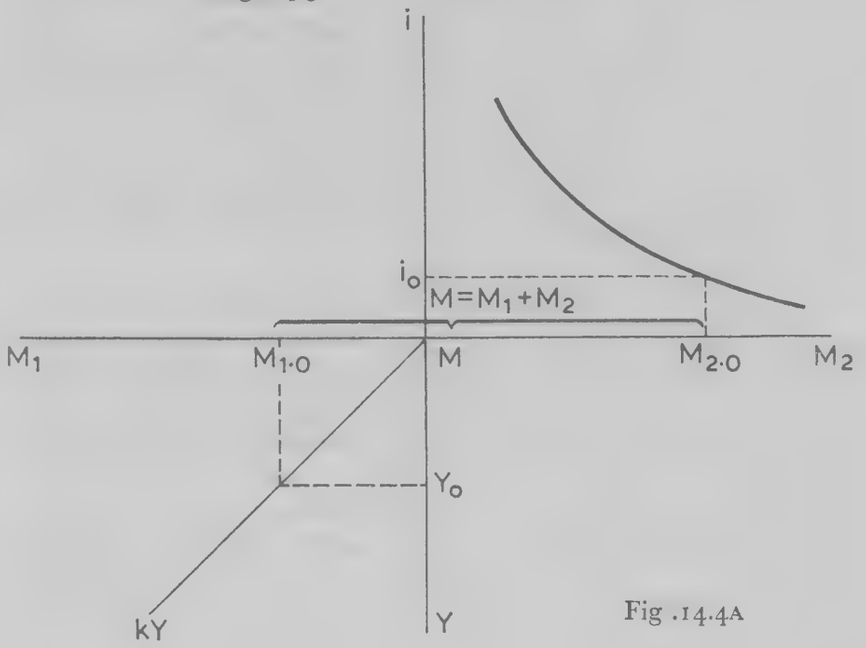


Fig .14.4A

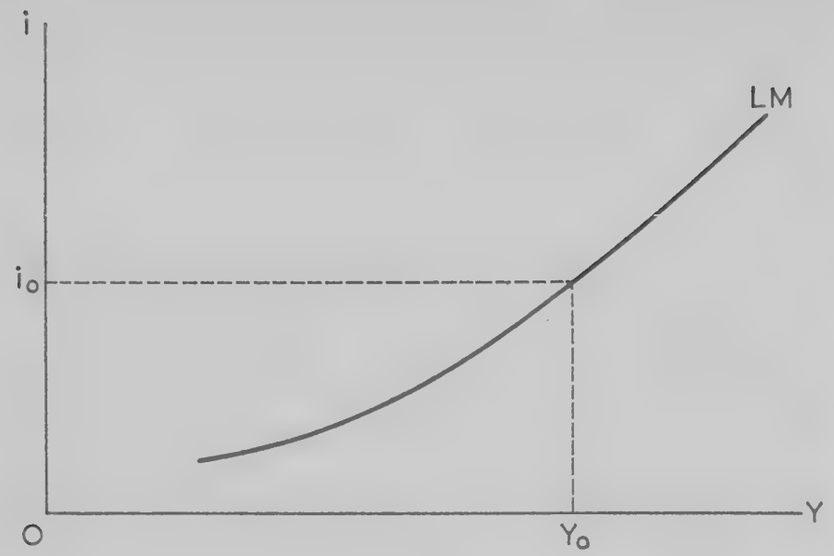


Fig. 14.4B

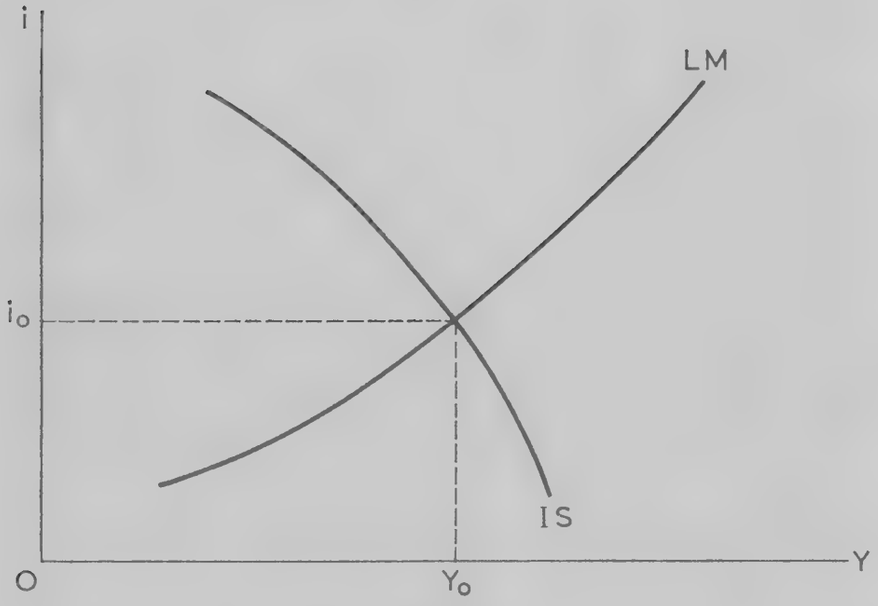


Fig. 14.5

Note

1. J. R. Hicks, 'Mr. Keynes and the "Classics": A Suggested Interpretation', *Econometrica*, vol. 5 (1937) 147–69, and in American Economic Association, *Readings in the Theory of Income Distribution* (London, 1946).

FULL EMPLOYMENT

Joan Robinson

FULL employment is bound to be a vague conception. First there is a quite arbitrary element in the hours per week that constitute full time. Does the opportunity to do occasional overtime come in? and if so, how much? Then there is the question of the number of bodies constituting the available labour force at any moment. Married women, students, wealthy rentiers – who is and who is not part of the labour force? Then there is the whole problem of self-employment. The term 'disguised unemployment' was originally invented to cover the case of the match-seller in the Strand who appeared during the slump. It has been extended to cover peasants with too small a holding to keep their families productively busy. But how productive? Unemployment shades over into low output per head. None of this matters for positive analysis. The terms can be defined for each problem in the manner appropriate to the matter to be discussed and need be given no more precision than the question requires.

For an ideological slogan this vagueness will not do. Full Employment is a Good Thing, and it is conceived to be attainable by wise policy. It is a blessed state, like equilibrium. We must be able to say what it is.

In his original definition, Keynes distinguishes voluntary unemployment, which may be due to the 'withdrawal of their labour by a body of workers because they do not choose to work for less than a certain real reward' and involuntary unemployment:

> Men are involuntarily unemployed if, in the event of a small rise in the price of wage-goods relatively to the money-wage, both the aggregate supply of labour willing to work for the current money-wage and the aggregate demand for it at that

wage would be greater than the existing volume of employment.[1]

Again:

> We have full employment when output has risen to a level at which the marginal return from a representative unit of the factors of production has fallen to the minimum figure at which a quantity of the factors sufficient to produce this output is available.

It is the 'marginal disutility of labour' that 'sets an upper limit' to the potential output.

The disutility of labour is a piece of Marshallian luggage that Keynes thoughtlessly carried with him.

Marshall describes a boy picking blackberries, who goes on until the *marginal utility* of another berry would not be sufficient to repay the *marginal disutility* of the extra effort.[2] In short, he goes on until he feels inclined to stop.

For a worker in a world without social insurance, who has the choice between taking a job at the going wage or having no wage at all, this conception is very much off the mark. Perhaps the reason that Marshall failed to notice its absurdity is connected with the peculiar system of remuneration of Oxford and Cambridge dons. A Fellow gets his college dividend quite independently of the number of units of effort that he puts forth, and he can also take pupils at so much a head. We all know how the marginal disutility of pupils increases with their number, and, with a basic income independently given, differences in the marginal utility of income may be assumed to be less important.

To a college Fellow it is an attractive idea to separate the marginal utility of income from the disutility of work experimentally. Pay him a larger and larger basic salary, and then see how the rate per head required to induce him to take a given number of pupils has to be varied. But unfortunately it would not really be scientific. The value of leisure is not independent of disposable purchasing power. This disutility of work may actually be negative if the alternative is nothing to do and nowhere to go, and is very high when the alternative is delightful, expensive treats.

Any measure that can be proposed for disutility of work will turn out to be elastic. Of all the concepts in the neo-classical bag this is the most irremediably metaphysical. Keynes did not attach any importance to it and was quite ready to accept as an alternative a simple-minded definition – that there is full employment when everyone who wants to has a job. But this is an unattainable upper limit. From an ideological standpoint it will not do to say that Full Employment can never be achieved.

Beveridge proposed the criterion of the relation between the number of unfilled vacancies and of registered unemployed. Both figures are obviously very rough indicators of what they are intended to indicate and, even if they were quite exact, an overall equality between them would not represent a critical point in the relation of supply and demand for labour, since the very coincidence of unfilled vacancies and unemployed workers shows that they do not fit, either because they are geographically separated or because the vacancies are for particular types of work which the unemployed cannot offer. A growing or falling excess of vacancies over unemployed is quite a useful indication, over the short run, of the movements of demand, and a drop in both would presumably indicate an improvement in the general conditions of mobility of labour or versatility of management. But there can be no virtue in taking an exact balance between them to indicate 'Full Employment' with capital letters.

(To digress for a moment, it is remarkable that when Beveridge was writing *Full Employment in a Free Society* in consultation with a number of young Keynesians, an *average* of unemployment of 3 per cent seemed quite a daring objective to propose. The idea that for more than twelve years we should not touch that figure, indeed that 2 per cent should come to be considered dangerously high, would at that time have seemed extravagant wishful thinking.)

From the first it was obvious that if we ever reached and maintained a low level of unemployment, with the same institutions of free wage bargaining and the same code of proper behaviour for trade unions that then obtained, the vicious spiral of rising prices, wages, prices would become chronic. Already at that time it could be argued that 'the point of full employment, so far from being an equilibrium resting place, appears to be a precipice over which, once it has reached the edge, the value of money must plunge into a bottomless abyss'.[3] This has turned out to be sadly true, and it is very troublesome ideologically, for both Full Employment and stable prices are Good Things. The solution sometimes found is to

say that when wages are rising there is *overfull* employment and to define Full Employment so as to include enough unemployment to prevent money-wage rates from rising faster than productivity. This is generally accompanied by the arbitrary assumption that some definite figure, say 3 per cent of unemployment, would keep prices stable and by the suggestion that this is a right and proper policy to aim at maintaining the postulated level.

Michal Kalecki, who discovered the *General Theory* independently, drew from it less optimistic conclusions than Keynes. When during the war it became clear that the new theory was firmly established and that the old trade cycle could be overcome, he predicted that we should find ourselves living under a political trade cycle:

> In the slump, either under the pressure of the masses, or even without it, public investment financed by borrowing will be undertaken to prevent large-scale unemployment. But if attempts are made to apply this method in order to maintain the high level of employment reached in the subsequent boom a strong opposition of 'business leaders' is likely to be encountered. As has already been argued, lasting full employment is not at all to their liking. The workers would 'get out of hand' and the 'captains of industry' would be anxious to 'teach them a lesson'. Moreover, the price increase in the upswing is to the disadvantage of small and big rentiers and makes them 'boom tired'.
>
> In this situation a powerful block is likely to be formed between big business and the *rentier* interests, and they would probably find more than one economist to declare that the situation was manifestly unsound. The pressure of all these forces, and in particular of big business – as a rule influential in Government departments – would most probably induce the Government to return to the orthodox policy of cutting down the budget deficit. A slump would follow in which Government spending policy would come again into its own. . . .
>
> The régime of the 'political business cycle' would be an artificial restoration of the position as it existed in nineteenth-century capitalism. Full employment would be reached only at the top of the boom, but slumps would be relatively mild and short-lived.[4]

Perhaps more emphasis should have been placed on the City than on big business, and more emphasis on monetary policy than budget deficits, but on the whole the above has proved to have been pretty near the mark.

All the same, the objection of low unemployment has turned out to be relatively weak (at least in Great Britain); certainly any return to heavy unemployment would be violently resisted. Taking it by and large, Full Employment has become an orthodox objective of policy.

The notion that Full Employment is attainable has become, as Keynes in some moods intended it to be, the new defence of *laisser-faire*. It is only necessary to remove one glaring defect from the private-enterprise system and it becomes once more an ideal.

Full Employment (with some reservations about not allowing it to get *overfull*) has become an aim of Conservative policy and the strongest argument against socialist critics. 'You used to complain, we now admit with some justification, that a capitalist system that permits heavy and chronic unemployment is indefensible. Now we offer you capitalism with a high and stable level of employment. You have nothing to complain of.'

Marxist critics have understood that Keynes's theory leads to conclusions which from their point of view are reactionary. They therefore deny the logic of his analysis and even find themselves in alliance with the protagonists of the humbug of finance which Keynes first attacked. For instance, Professor Baran is not content with showing that an economic system that can maintain prosperity only by expenditure on armaments is a menace to humanity, morally abhorrent and politically disreputable; he also has to bring in the Quantity Theory of Money to show that it cannot work because Government expenditure causes inflation.[5]

This is another example of confusion between logic and ideology. Because Keynes has shown a way for the capitalist system to remove its most obvious defect, he is a reactionary and therefore his theory is false.

But if his theory were false it would be quite harmless. Just because the diagnosis was correct, the treatment is found to work and the life of the patient is being prolonged, disconcerting his would-be heirs.

The reason why Full Employment has become a right-wing

slogan is that if employment is an end in itself no questions can be asked about its content. What is work for? Only to keep the workers out of mischief. Any product is as good as any other.

Keynes fired off his paradoxes to penetrate the thick walls of obscurantism of the old *laisser-faire* orthodoxy:

> Ancient Egypt was doubly fortunate, and doubtless owed to this its fabled wealth, in that it possessed *two* activities, namely, pyramid-building as well as the search for the precious metals, the fruits of which, since they could not serve the needs of man by being consumed, did not stale with abundance. The Middle Ages built cathedrals and sang dirges. Two pyramids, two masses for the dead, are twice as good as one; but not so two railways from London to York. Thus we are so sensible, have schooled ourselves to so close a semblance of prudent financiers, taking careful thought before we add to the 'financial' burdens of posterity by building them houses to live in, that we have no such easy escape from the sufferings of unemployment.

And he argues in favour of waste when no useful outlet for investment is seen to be profitable:

> In so far as millionaires find their satisfaction in building mighty mansions to contain their bodies when alive and pyramids to shelter them after death, or, repenting of their sins, erect cathedrals and endow monasteries or foreign missions, the day when abundance of capital will interfere with abundance of output may be postponed. 'To dig holes in the ground', paid for out of savings, will increase, not only employment, but the real national dividend of useful goods and services.

But he adds:

> It is not reasonable, however, that a sensible community should be content to remain dependent on such fortuitous and often wasteful mitigations when once we understand the influences upon which effective demand depends.

Nowadays the paradoxes are taken in sober earnest and building weapons that become obsolete faster than they can be constructed has turned out far better than pyramids ever did to keep up profit without adding to wealth. The relapse on Wall Street that follows any symptom of relaxation in the Cold War is a clear demonstration of the correctness of Keynes's theory, but also a demonstration of the falsity of his optimistic view that, when the theory was understood, reason would prevail.

He was himself partly to blame for the perversion of his ideas, for he failed to see that, once the principle has been established that maintaining employment is a public concern, the question of what employment shall be for becomes a political issue.

In the last chapter of the *General Theory* he falls into the fallacy of supposing that there is some kind of *neutral* policy that a Government can pursue, to maintain effective demand in general, without having any influence upon any particular demand for anything. The Government has to undertake 'the task of adjusting to one another the propensity to consume and the inducement to invest', but everything else is best left to 'the free play of economic forces'.

This is a metaphysical conception as unseizable as *abstract labour* or *total utility*. What is a policy which *merely* adjusts the demand for investable resources to the supply?

To increase effective demand when it threatens to flag, various means can be used: to reduce taxation or to shift the burden from those most likely to increase their consumption to those most likely to reduce their savings; to foster competition so as to reduce profit margins; to increase subsidies or outlays on social services – all means which tend to reduce inequalities in consumption. Or Government expenditure on investment can be increased, directly or through nationalised industries, or reductions in taxation and credit policy can be used to encourage private investment. Contrariwise, when effective demand seems excessive, taxes to discourage consumption, credit restriction and reduced Government expenditure can be brought into play. And all this has to be worked out so as to preserve the balance of trade at some level or other, as well as to preserve employment. What is a *neutral* policy? What mixture of these means is it that leaves private enterprise unaffected in content and acts only on the quantity?

There is in some quarters a great affection for credit policy because it seems the least selective and somehow lives up to the ideal of a single overall neutral regulation of the economy. The enormous ideological attraction of the Quantity Theory of Money,

that kept it going for nearly forty years after its logical content was exploded, is due to the fact that it conceals the problem of political choice under an apparently impersonal mechanism.

Recent experiments have shown, however, that there is no such thing as a purely quantitative, overall financial policy. We are fortunate in having had an official report that finally discards the old mumbo-jumbo. But just because the Radcliffe Committee is clear on this point it has no definite recommendations to make. There is no simple right policy; it is all a matter of judgement.

The Keynesian revolution has destroyed the old soporific doctrines, and its own metaphysics is thin and easy to see through. We are left in the uncomfortable situation of having to think for ourselves.

Notes

1. J. M. Keynes, *General Theory of Employment, Interest and Money* (London, 1936).
2. A. Marshall, *Principles of Economics* (London, 9th edn, 1961).
3. Joan Robinson, *Essays in the Theory of Employment* (London, 1937).
4. M. Kalecki, 'Political Aspects of Full Employment', *Political Quarterly*, Oct.–Dec. 1943.
5. P. Baran, *The Political Economy of Growth* (New York, 1957).

In a developed 'market economy', commodities are not exchanged for one another by direct barter, but indirectly through the medium of money. Because the use of money extends through the whole economy, it is important to ensure that the existing quantity is neither too small nor too large to do the work required of it, and questions of monetary policy have often been the subject of acute controversy. In the following extract, Sir John Hicks traces the development of economists' views on money and its connection with the working of the economy as a whole.

MONETARY THEORY AND HISTORY: AN ATTEMPT AT PERSPECTIVE

J. R. Hicks

I HAVE attempted, on two previous occasions, to elucidate the relation between Keynes and those whom he called 'classics'. The method that was employed in those papers was analytical; analysis was one of the things that needed to be done; I believe that the analytical method, up to a point, did justify itself. It has nevertheless left me in some ways dissatisfied. The question is not merely analytical, it is also historical. What did the pre-Keynesian writers say, and why did they say it? In order to complete the discussion, there are some further matters in this direction which need to be explored.

'Classics', as used by Keynes, was a confusing description. Before Keynes (and even since Keynes, by those who are not specially concerned with Keynesian controversy), 'classical economics' has been used in a different and more restricted sense. What we otherwise mean by classical economics is the economics of Adam Smith, Ricardo and their contemporaries; not the economics of those whom we have now had to call the 'neo-classics' – those who flourished in the three-quarters of a century before 1936. Keynes certainly means his 'classics' to include the neo-classics; it may indeed be some of the latter whom his cap most exactly fits. But I think that he spoke of them all as 'classics' because he perceived (quite rightly perceived) that some of the things he was attacking came down from Smith and Ricardo, especially perhaps Ricardo, whom he was right in identifying as the chief originator of what he called the classical tradition. It was indeed the old classical period that was the formative period. But if one goes back to that period it is not just this 'classical' tradition that is forming; other things were happening as well. There were differences, or qualifications; I believe that we understand much better what happened if we give some weight to them. For they do not only show how there came to be an 'orthodoxy' for Keynes to attack; they also throw light upon the Keynesian orthodoxy which has so largely replaced it. It is very desirable, at the stage we have now reached (and even, as we shall see, in relation to our own contemporary problems), that we should make some effort to get the whole story into proper focus.

I

Monetary theory is less abstract than most economic theory; it cannot avoid a relation to reality, which in other economic theory is sometimes missing. It belongs to monetary history, in a way that economic theory does not always belong to economic history. Indeed it does so in two ways which need to be distinguished.

It is noticeable, on the one hand, that a large part of the best work on money is topical. It has been prompted by particular episodes, by particular experiences of the writer's own time. All theorising is simplifying, cutting out the unimportant and leaving what is thought to be important, in the hope that by simplifying we may increase understanding. Sometimes what is sought is a general understanding; but with monetary theory it is more often a particular understanding – an understanding directed towards a particular problem, normally a problem of the time at which the work in question is written. So monetary theories arise out of monetary disturbances. This is obviously true of the *General Theory*, which is the book of the great depression – the world depression – of the nineteen-thirties; it is also true of Keynes's

other version, the *Treatise on Money*, which differs from the *General Theory* quite largely because it is directed at a different contemporary problem. Though the *Treatise* was published in 1930, after the depression had begun, it must largely have been written earlier. Its world is not the world of the depression, it is the world of the restored gold standard. Its problem is how the restored gold standard is to be made to work. Now much the same is true of Ricardo and his contemporaries, a century earlier.

Ricardo's monetary writings (the earliest is dated 1808, and he died in 1823) cover a period of war inflation, in the last stages of the British war against Napoleon, and a period of reconstruction, and attempted stabilisation, after the peace. His problems are the problems of those years. Thus his work, like Keynes's, was the result of a challenge – a challenge from contemporary experience. It is possible that one reason why monetary theory did get a bit ossified in its neo-classical phase is that in that phase there seemed to be no similar challenge. At least in the seventies and eighties, when Marshall (the greatest of the neo-classicals) could have responded to a challenge, it was just not there; you could not get much of a kick out of bimetallism! You cannot get brilliant answers to a dull question-paper; and the question-paper that was set to Marshall by his monetary facts really was a bit dull.

That topicality is one way in which monetary theory is historically conditioned; but there is another also.

Throughout the whole time – back before Ricardo, forward after Keynes – money itself has been evolving. The change from metallic money to paper money is obvious; but there are other things which have gone with that change, of even greater importance, which are not so easy to recognise and to assess. Even if we say that metallic money has given place to credit money, we are still not getting to the bottom of what has happened. For credit money is just a part of a whole credit structure that extends outside money; it is closely interwoven with a whole system of debts and credits, of claims and obligations, some of which are money, some of which are not, and some of which are on the edge of being money. The obvious change in the money medium, from 'full-bodied' coins to notes and bank deposits, is just a part of a wider development, the development of a financial system. This has taken the form of the growth of financial institutions, not just banks, but other 'financial intermediaries' as well; it has carried with it a fundamental change in the financial activities of governments. In the course of these changes there has been a change in the whole character of the monetary system. In a world of banks and insurance companies, money markets and stock exchanges, money is quite a different thing from what it was before these institutions came into being.

This evolution has been going on ever since the time of Ricardo (its beginnings, of course, are much earlier); it clearly called, as it proceeded, for a radical revision of monetary theory. As the actual system changed, the theoretical simplification ought to have changed with it. We can now see that it did not change sufficiently; there was a lag. But the reason for the lag was not just laziness or sleepiness; there was an obstacle to be overcome.

On the theoretical level – in terms of basic principles – the evolution that was occurring had two aspects. From one of them it was a natural piece of economising. Metallic money is an expensive way of performing a simple function; why waste resources in digging up gold from the ground when pieces of paper (or mere book entries) which can be provided, and transported, at a fraction of the cost will do as well? That is the reason why the credit system grows: that it provides a medium of exchange at much lower cost. But on the other side there is the penalty that the credit system is an unstable system. It rests upon confidence and trust; when trust is absent it can just shrivel up. It is unstable in the other direction too; when there is too much 'confidence' or optimism it can explode in bursts of speculation. Thus in order for a credit system to work smoothly, it needs an institutional framework which shall restrain it on the one hand, and shall support it on the other. To find a framework which can be relied on to give support when it is needed, and to impose restraint just when it is needed, is very difficult; I do not think it has ever been perfectly solved. Even in this day we do not really know the answer.

II

When Ricardo and his contemporaries saw this problem (some of them, as we shall see, did certainly begin to see it) it frightened them quite a bit. So they tried to hack their way out.

At their date, it must be remembered, the evolution of the credit system had not gone very far. Thus it was natural still to regard the

metallic money as primary; the notes and bills (which already existed) as a tiresome, but secondary, qualification. If only the secondary money would behave like primary money, there would be no trouble! So let us try to make it behave like primary money. Then we can carry on in our thinking with a simple, easily understandable, primary money model.

That, in effect, is what Ricardo did. But it is not enough to recognise that his model is a metallic money model. We should still ask: as a metallic money model, for a system without developed credit, was it right? Would it be true, even in a world where all borrowing and lending was long-term borrowing and lending (for that, at the least, must be assumed if we are to have no credit), that interest rates will be entirely determined by saving and investment, that the level of activity will be solely determined by the real factors in the system, and that the quantity of money will solely act upon the level of prices? These are the 'classical' doctrines, which we associate with Ricardo. It is evident that Ricardo relied upon them as an approximation to reality; but I wonder whether even he, even for his metallic money model, thought that they were *exactly* true.

For there are writings, serious and intelligent writings, which were available to Ricardo (which indeed we can be quite sure that he must have read), and in which some at least of the qualifications are quite clearly set out. The most striking of these is the essay 'Of Money' by David Hume. Hume, of course, is most famous as a philosopher; but he was no mean economist also. He was a close friend of Adam Smith, but his work on economics is earlier than Smith's; it was Smith who learned from Hume, not the other way about. For Hume, writing in 1750, it was natural to assume a purely metallic money; for the growth of the credit system, though it had begun, was in his day at a much earlier stage than it was in Ricardo's. But his analysis of the working of a monetary system is not simply what Keynes would have called 'classical'. For Hume the Quantity Theory of Money is *an equilibrium condition.*

> Money [he says] is nothing but the representation of labour and commodities, and serves only as a method of rating or estimating them. Where coin is in greater plenty – as a greater quantity of it is required to represent the same quantity of goods – it can have no effect, either good or bad, taking a nation within itself; any more than it would make an alteration in a merchant's books, if instead of the Arabian method of notation, which requires few characters, he should make use of the Roman, which requires a great many.

That, for Hume, is the main point; the *classical,* Quantity Theory, point. But Hume goes on:

> Notwithstanding this conclusion, which must be allowed just, it is certain that since the discovery of the mines in America, industry has increased in all the nations of Europe, except in the possessors of those mines; and this may justly be ascribed, among other reasons, to the increase of gold and silver. Accordingly we find that in every kingdom, into which money begins to flow in greater abundance than formerly, everything takes on a new face; labour and industry gain life; the merchant becomes more enterprising, and even the farmer follows his plough with greater alacrity and attention. . . .
>
> To account then for this phenomenon, we must consider, that though the high price of commodities be a necessary consequence of the increase of gold and silver, yet it follows not immediately upon that increase; but some time is required before the money circulates through the whole state, and makes its effect be felt on all ranks of people. At first, no alteration is perceived; by degrees the price rises, first of one commodity, then of another; till the whole at last reaches a just proportion with the new quantity of specie which is in the kingdom. In my opinion, it is only in this interval or intermediate situation, between the acquisition of money and the rise of prices, that the increasing quantity of gold and silver is favourable to industry.

Putting the point in our language, the Quantity Theory is valid as a long-term equilibrium condition; but in the short period, while the supply of money is increasing, the increase can be a real stimulus. Now the interesting question is: if Ricardo had been challenged, and had been held to it, would he – could he – have said anything different?

There is one piece of evidence which suggests that after all he would not. It is evidently implied by what Hume says about the stimulus of monetary expansion that the effect of monetary

contraction will be the reverse. If the supply of money is reduced, or if money is taken out of circulation by hoarding, prices will fall; but while they are falling, business will be depressed. Now though Ricardo (so far as I know) never admitted this as a principle, he acted as if he believed it. For though he was prepared to accept some fall in prices, after the war (through which he had been living for so long) was over, in order to achieve what he thought was the great good of monetary stability, he was very much concerned to ensure that the fall in prices should not be unnecessarily large. That was the purpose of his famous 'Ingot Plan' which was designed to prevent an excessive fall in the money in circulation. I do not see why he should have been so bothered upon this, if he regarded money prices – the course of money prices over time – as having no *real* importance.

Quite a number of things will fit into place if we suppose that the classical economists, of this important and in so many ways constructive period, did have some such short-period theory, somewhere at the back of their minds, though they preferred not to emphasise it. One can see several reasons why they should have preferred not to emphasise it.

There was a special reason in Ricardo's own case, that he was the great creator of the static equilibrium method in economics; he was showing, for the first time, how much could be done with that method; it was his method, and he was reluctant to turn away from it. Not only here, but in many other applications, he tended to rush from one equilibrium to another much too quick. Then there is the point, which I have already mentioned, that the monetary theory of that time was oriented towards a particular practical problem, the restoration of a particular kind of stability, a problem that seemed eminently suitable for analysis in equilibrium terms. But there was also another reason, the most important of all. They were terribly afraid that if too much weight were given to short-period effects, it would play into the hands of crude inflationists. The long period, it would be said, is just a succession of short periods. Why not keep the stimulus going, when the first dose is exhausted, by another dose? They were afraid of that question, for they did not know the answer to it. Yet they felt in their bones that the suggestion in it was wrong.

Nowadays, I think, we know the answer. We know it in theory, and we have seen it confirmed in practice. Inflation does give a stimulus, but the stimulus is greatest when the inflation starts – when it starts from a condition that has been non-inflationary. If the inflation continues, people get adjusted to it. But when people are adjusted to it, when they *expect* rising prices, the mere occurrence of what has been expected is no longer stimulating. Nor can the fade-out be prevented by accelerating the inflation; for acceleration of inflation can be expected too. It is perfectly possible to have an 'inflationary equilibrium', in which prices go on rising, even for years, more or less as they are expected to rise; but then there is no stimulus. In real terms, in terms of production and even of employment, the economy may be very depressed. We have seen examples of this in our time, in South America and in East Asia, in Argentina and in Indonesia, for instance; they are not pretty. The classical economists were quite right in refusing to look that way, though they did not quite know just why they were refusing.

You may ask for some evidence that this really was their state of mind. I can say that I know one very eminent nineteenth-century economist in whom the statement of it is almost explicit – so nearly explicit as to be readily recognisable; this is John Stuart Mill. There is no doubt at all that Mill did have a short-period theory such as I have been describing. He has written it out for us himself, in his essay on the 'Influence of Consumption upon Production' (in his *Essays on Unsettled Questions*). There is no doubt at all that when Mill wrote that essay he was on what we have been calling Hume's side. But it is a remarkable thing that when we turn to Mill's *Principles* (1848 – the *Essays* were published in 1844, but are stated to have been written earlier), we find no reference at all to the argument of the essay. In the *Principles* (see especially the chapter on 'Excess of Supply') Mill appears, on all this side, to be just a hard-boiled 'classic'. The argument of the essay is not withdrawn, but it is just not there. I feel sure that we can best explain Mill's position by supposing that he always held to what he had said in the essay, but he did not want to emphasise it, for he held that it was *dangerous*.

III

All this can be said before we come (at all seriously) to the evolution of credit. But in fact, even in Ricardo's time – and still

more by the time of Mill, a generation later – the development of credit had gone quite far. In trying to treat the monetary system as if it was a metallic system, or could be forced into the mould of a metallic system, Ricardo was looking backward. The monetary system in terms of which he was thinking was already, in his day, a thing of the past.

There was at least one of his contemporaries to whom this was quite familiar. Ricardo must certainly have read the 'Enquiry into the Nature and Effects of the Paper Credit of Great Britain', by Henry Thornton (1802 – six years earlier than any of Ricardo's writings); but he passed it by. Thornton has a short-period theory, not unlike Hume's; but it is worked out in terms of a credit economy. Thornton's is in fact the best analysis of the working of a credit system which was given by any of the older economists. As a short-period theory, it can be rewritten in a form which brings it quite close to Keynes.

For Thornton does not only recognise (as Hume had done) that in the short period monetary causes may have real effects; he also recognised that in a credit system the reverse can happen. Real causes have monetary effects. A credit system will expand automatically if there are real causes making for an expansion in activity (more or less what Keynes was later to call a rise in the marginal efficiency of capital); it will contract automatically if there is a panic (involving a sharp rise in liquidity preference), or if there are changes in the demand for capital which make for contraction. Some of these expansions and contractions are desirable, some highly undesirable. Thornton accordingly held that a credit system must be *managed*. It must be managed by a Central Bank, whose operations must be determined by judgement, and cannot be reduced to procedure by a mechanical rule.

The objectives to which Thornton thought this management should be directed were quite complex; it is most interesting to find that the avoidance of unnecessary unemployment does find a place among them. But the weighting which he gives to different objectives is not unnaturally very different from the weighting which would have been given by Keynes. I need not go into that; the thing on which I want to insist is that Thornton does believe in the necessity of monetary management; on that crucial matter he is on the same side as Keynes – on the opposite side from Ricardo.

Now where did Mill stand on that point? In the light of what I have already said about Mill, that is quite an interesting question. Mill was writing after Ricardo, and after Thornton; there is no doubt that he was acquainted with the work of both. (There is quite a long passage from Thornton's book that is quoted in full in Mill's *Principles*.) Mill was perfectly well aware that he was dealing with what had become a credit economy. Though he did not have Thornton's practical experience, he had good observers of the financial scene (especially Tooke and Fullarton) on whose work he could draw; and that a mighty change had occurred was by his time unmistakable. Ricardo had died in 1823; there followed in 1825 the first of that notable series of credit crises which mark the economic history of nineteenth-century England (1825, 1839, 1847, 1857, 1866) – demonstrations of the instability of credit which could not be overlooked. This period of credit crises (much more distinctive as credit crises than some of those which have marked later 'trade cycles') was the period during which Mill was writing. He had to take a line about them.

The analysis which Mill gives of the working of credit expansion is really rather good. He saw that it was likely to get started, not by an expansion of bank credit, but by an expansion of trade credit, without the banks (or the main banks) being at that stage much involved. As the boom develops, it requires to be fortified by more secure forms of credit, so the pressure is carried back from the circumference to the centre of the banking system. It is essential, at that point, that the centre should hold firm; it must protect itself, but only in order to be able to spread security around it. Because he understands that this is the way the system works, Mill is unable to follow Ricardo in looking for mechanical rules by which credit is to be controlled; it can only be controlled by quite subtle appreciation of the 'feel' of the market, by monetary *policy*. In all this Mill is on the same side as Thornton; he believes in monetary management.

Yet Mill's idea of monetary management seems oddly one-sided. He is concerned to prevent booms; but he seems much less concerned to prevent slumps. One must clearly not attribute this to any lack of sensitivity to the suffering caused by slumps; for Mill is outstanding among the older economists for his social sympathy. It does, however, seem to be explicable in terms of his particular experience. The slumps of his time followed on directly

from the crises; they seemed obviously traceable to the disorganisation caused by the crisis. Once that disorganisation was set right, trade recovered. It was natural to conclude that if the booms could be prevented, if credit could be prevented from over-expanding, the slumps would be prevented too. That is a point (of course) on which a Keynesian economist would refuse to follow him. Mill, I think, just took it for granted; for in his time things did look just like that. It did not seem to be necessary to inquire further.

IV

I can now conclude what I have to say about these old classical economists. It is an old story; but I have tried to show that it is a significant story. There is much more to it than the formation of an orthodoxy. The classical economists did not all say the same thing. There were differences between them, and some of the differences which then came up for the first time are still important.

There were, at the least, two strands in classical economics. There was one (represented, roughly speaking, by Ricardo and his followers) which maintained that all would be well if by some device credit money could be made to behave like metallic money; there was another (represented, so far as I have taken the story, by Thornton and Mill) which held that credit money must be managed, even though (as was admitted) it is difficult to manage it. This is a major difference, and it has outlasted Keynes. In this sense we still have Ricardians among us. When Milton Friedman tells us that we should have 'a legislated rule instructing the monetary authority to achieve a specified rate of growth in the stock of money', he is being Ricardian. When Jacques Rueff, in many current speeches, tries to push us into revaluing gold, in order to replace credit money (if only in international transactions) by gold transfers, he is being Ricardian. But this Ricardian strand is only one strand in 'classical' monetary theory; we have Thornton and Mill to prove the contrary.

What happened to these two strands, after the time of Mill, in the neo-classical phase? I must here be very brief. What mainly happened, I think, was this. It was the Ricardian which remained the official doctrine; or perhaps one should say that many things which only make sense in Ricardian terms remained official doctrine. It survived quite sufficiently to be presentable as *the* classical theory for Keynes to attack. Bankers talked Ricardo in their speeches. Descriptions of the monetary system were given in textbooks – in terms of cash ratios, fiduciary issues and so on – by which the quite fully developed credit system which then existed was translated into Ricardian terms. That, however, was theory; in practice it was the Thornton–Mill school which won out. On strict Ricardian principles, there should have been no need for Central Banks. A Currency Board, working on a rule, should have been enough; but in fact, during this period, there was a growth of Central Banks. There was quite odd double-talk about it. The Bank of England itself was supposed to be Ricardianised by the Bank Charter Act of 1844, which divided it into an Issue Department, that was just to be a Currency Board, and a Banking Department that was to differ little from an ordinary bank. But that was not in fact what happened. The Banking Department became a Central Bank, which, at least to a limited extent, did exercise monetary policy, It had to be cautious about it, since it was not in accordance with official doctrine that it should do so. I think nevertheless that there is no doubt that that is what it did.

This was indeed an odd situation; why did it persist? I think it suited the bankers very well to represent themselves, and even to think of themselves, as passive – just keeping to the 'rules'. For if it had been suspected that they were actively controlling, people would have asked: what right have they to control? Their actions are affecting everyone, affecting everyone (on occasion) very deeply. What right have they, who have not come up through the regular channels of democratic government, to arrogate to themselves such power? So it was useful to them to keep a screen in front of them. Once it was suspected that they were exercising control, their right to control was bound to be called in question. And so, in the end, it proved.

V

That anti-banker revolution was one aspect of the 'Keynesian revolution' as it worked out in practice; but it was mixed up with the other aspect, concerning the kind of control that should be exercised, not who should exercise it. To this I must now turn.

The kind of banking control that Mill envisaged was directed against excessive expansion, not the other way about. I have suggested that in leaning that way, Mill may have been influenced by his particular experience, by the conditions of the time when he lived; but when we come to Keynes, is there not something of a similar possibility? The *General Theory* was written in the world depression; even the twenties, when Keynes was writing the *Treatise*, were in England years of semi-slump, or of a boom that misfired. It was natural for Keynes, writing when he did, to take as his base a depressed economy. His practical problem was emergence from depression; he looked at the world from a depression point of view. I am myself quite sufficiently Keynesian to be convinced that the Mill kind of control (control which can be exercised by a Central Bank, if it is allowed to exercise it) is by no means enough. But the picture one so easily gets from Keynes, of an economy which will always be under-functioning, unless it is boosted by deliberate expansionary policy, seems to me to be just as one-sided. Ideas which arose in the thirties, and which were appropriate in the thirties, have been carried over, through Keynes's influence, into a world where they do not at all so obviously belong.

However that may be, we have unquestionably learned from Keynes that control through banking is one-sided. It can be effective against inflation and over-expansion; against over-contraction it is relatively powerless. Banks can restrict expansion by refusing to lend; but they cannot force expansion just by offering to lend, on whatever easy terms. It can be that business is feeling so dismal that even on the most favourable possible terms (which are consistent with the banking system making any sort of a profit) loans will not be taken up. (Even so, it is possible for a follower of Mill to retort, as I think Dennis Robertson might have done, more or less, that this is a very odd state of mind for businessmen to be in; it must be due to some cause. It must be due to some obstruction in the business circulation, surplus stocks that can and will be worked off, bad debts that can and will be settled; the most that is needed is some help on the way to this natural recovery.) Keynes has nevertheless clearly shown that a depressed economy can be boosted by government spending; and once that is granted, the way to the acceptance of government spending as a *general* booster is hard to resist.

Two general reflections about the post-Keynesian world, which seem to arise out of what I have been saying, may be made in conclusion.

(1) There was a first phase – I think one can now say that it was just a first phase – in which it seemed to be a simple matter for the whole control of economic activity to be taken over by the government budget. This was the heyday of 'fiscal policy'. If there was a tendency to contraction, more should be spent; if there was a tendency to over-expansion, taxation should be raised. For a while this did not work badly; but it was an arrangement that had its defects. When the principle is so crudely stated, the defects are rather obvious.

It is easy enough for Government to spend more; that is a simple and popular policy. Some sorts of taxes (taxes that will be paid by people whose votes you don't want) are also popular. But the supply of such convenient taxes tends to give out; control by unpopular taxes is a different matter. When the Keynesian prescription calls upon the Minister to impose unpopular taxes, he begins to long for the old days, when he had a banker behind whom he could hide. 'This business of monetary control', he will be saying, 'is a technical matter. If I deal with it by fiscal policy, though I can say that I am taking the advice of my experts (my economic advisers and so on), I have to take the responsibility myself for these nasty decisions. It would be much better to let the experts do the job themselves.' So he makes passes to hand it back to the bankers. But he cannot really hand it back to the bankers, for everyone knows that he controls the bankers. They are no longer in a position to carry out an independent monetary policy, as they could do – to some limited extent – in former times. What happens is that the control gets lost between them. The control – the Keynesian control – cannot be properly exercised until there is a government that is strong enough to take unpopular measures, and to assume responsibility for them.

In this last passage, I (in my turn) have no doubt been influenced – perhaps over-influenced – by recent experience in Britain. Still I think that the situation is one that can occur elsewhere, and may indeed have occurred elsewhere on some occasions. It is asking quite a lot of a government that it should continue, year in and year out, to maintain a Keynesian control, in both directions.

(2) The other thing on which I want to say something is the international side. This, in its turn, splits up into two parts. In one of them we are looking outwards, from the point of view of the single nation, at its international economic relations. In the other we are looking at the well-being of the international economy as a whole. This latter may appear so remote as hardly to deserve attention; but it is not in fact to the advantage of any single nation that it should be forgotten, and left to look after itself.

In the old days, when monetary control was exercised by Central Bankers, their primary concern was the stability of the currency in terms of foreign exchange. That objective, in the new era, has been deposed from its old position. It was recognised, even in the initial Charter of the I.M.F., that stability of foreign exchange was not so sacred an object as Central Bankers had tried to make it in former times. Arrangements were made for agreed adjustment of exchange rates on the occasion of what was called a 'fundamental disequilibrium'. I would in no way question that this is indeed a real problem, a problem which the world has not (as yet) done much to solve. I would nevertheless maintain that it has been quite overshadowed by another, by the inflations (active and creeping inflations) which are traceable, not to 'fundamental disequilibrium', but to the implementation of pseudo-Keynesian policies by weak and irresolute governments. It is this which has been the main cause of exchange instability, and – what maybe is even worse – of feared exchange instability. There is widespread fear of exchange instability, just because, in the Keynesian epoch, the impression has got around that exchange depreciation, or devaluation, does not matter all that much.

I think it does matter. It may sometimes be necessary, as the least evil among a choice of evils. But it is an evil, which should not be regarded, as it is so often regarded at present, as a normal and natural way out.

The reason why it is an evil becomes apparent when we look at the matter from the other angle, that of the international economy. The story which I have been telling in terms of those old characters – Thornton and Mill and Ricardo – is repeating itself, in our day, on the international stage. There is the same movement from metallic money (gold) to credit money, going on there too. There is the same problem of the instability of credit. There is the same need that international credit should be managed, in order to be secure.

We do not need, on the international plane, to feel the Keynesian fear, that purely monetary management will be unable to fight depression; on the international plane it is not depression, in the old sense, that is the danger. National governments, taught by Keynes, however indirectly, can see to that. What is liable to happen, if there is a failure of international credit, is that nations will turn in upon themselves, becoming more autarkic or more protectionist, impoverishing themselves and each other by refusing to trade with each other. (And this means, we can already see, refusing to *aid* each other.) That is the danger with which we are confronted; we can already see that it is no imaginary danger at this present time.

The remedy, my old nineteenth-century experience would tell us, would be an International Central Bank, an International Bank which would underpin the credit structure, but in order to underpin it must have some control over it. That was what Keynes, who understood this international aspect very clearly, wanted to get at Bretton Woods; but all he got was a Currency Board (for it is little more than a Currency Board, being so tied up with rules and regulations) – the I.M.F. That, we are finding – and Mill could have told us, one hundred and twenty years ago, that it is what we should find – is not enough. But how should the powers, which governments have been unwilling to entrust to their own Central Banks (once they have realised what is involved), be entrusted to an International Bank? That is the dilemma, the old dilemma, to which we have now come back, on the international plane.

Stated like that, the problem looks insoluble. In such black and white terms, it probably is. But to set rules against no rules is to make too sharp an opposition. Can we find rules that are acceptable to national pride, and to national self-interest, and which yet give scope for some minimum of management – just enough to give the international credit structure the security it so sorely needs? It will be a narrow passage, but one must hope that there will be a way through.

The process of production generates incomes, which the Government taxes and partially redistributes through 'transfers' such as retirement pensions. The resulting pattern of personal incomes shows wide variations in the size of incomes, as described by R. J. Nicholson in the following article.

THE DISTRIBUTION OF PERSONAL INCOME

R. J. Nicholson

In 1957, Professor F. W. Paish contributed an article[1] to *Lloyds Bank Review* in which he examined the changes in the distribution of personal income over the years 1938 to 1955, and two years later Dr H. F. Lydall, in a contribution[2] to the *Journal of the Royal Statistical Society*, made a more detailed study of the changes over the longer period 1938 to 1957. These writers showed that the concentration of personal income, both before and after tax, in the hands of the top 5 per cent of income recipients was substantially reduced after the war, compared with 1938, and that this tendency towards greater equality continued during the post-war years which their analyses covered. Indeed, Lydall found that 'If anything the tendency towards reduced inequality of pre-tax income seems to have been accelerating.' The present article, linking up with this earlier work, examines the distribution of personal income in the period 1949–63 to see what light more recent data shed on the trends found for the earlier post-war years.

DEFINING PERSONAL INCOME

Both Paish and Lydall used the estimates of personal income as set out and tabulated in the national income and expenditure accounts for the U.K. These figures have been criticised, however, particularly by Professor Titmuss,[3] as being inadequate for studying changes in the distribution of incomes. A further analysis based on these data thus requires some preliminary justification.

First, it is argued that the number of incomes in the national income tabulations is a compromise between the number of family incomes and the number of individual incomes, since it is derived from Inland Revenue records in which the incomes of husband and wife are counted as one, and all other incomes are counted separately. Hence, the distribution of personal income could be affected by changes in the age of marriage and by changes in the proportion of the total population married and of married women going out to work. These changes may be important when pre-war and post-war distributions of income are compared, because considerable sociological movements took place as a result of the war, and it was in connection with such comparisons that Titmuss made his sharpest criticisms. However, these demographical and sociological changes have been less rapid since 1949 and would not distort comparisons of income distributions over the last decade.

Secondly, this definition of personal income is criticised on the ground that certain claims on wealth are excluded: for instance, undistributed profits and 'tax avoidance' incomes which fall outside income-tax regulations. The effect of extending the definition of personal income to include these other claims on wealth is by no means certain and there are, anyway, differences of opinion as to what the appropriate definition should be, e.g. whether companies' undistributed profits should be included. It is probable that the shape of the income distribution, i.e. the proportionate shares of different income groups, might be modified but it does not follow that the *trend of changes* in income distribution derived from the national income tables would be discredited. Indeed, even in pre-war/post-war comparisons, Lydall's estimates suggest that the trends would not be fundamentally altered. In the national income figures of personal income we do, at least, have a consistently defined aggregate which must account for a major proportion of total income on any definition and one which is adequate for picking out changing trends in the distribution of incomes in post-war years.

In the following tables, comparisons are made between various

levels of income and shares of particular groups in terms of incomes ranged in order of size. For example, the top 1 per cent of incomes all receive more than a certain amount, those in the top 5 per cent are all above a lower figure, those in the top 30 per cent are above a still smaller figure, and so on. Thus, Table 16.1 shows that the top 1 per cent of incomes were all above £1861 in 1949 and exceeded £3364 in 1963, while the top 40 per cent of incomes had a minimum of £304 and £773 in these two years. These money incomes which divide off the top 1 per cent, top 5 per cent, top 10 per cent and so on of incomes may conveniently be referred to as 'percentile incomes'.[4] Secondly, the distribution of incomes can be considered in terms of percentage shares. From Table 16.3, for example, it will be seen that in 1963 the top 1 per cent of incomes received 7·9 per cent of the total, the next 4 per cent received 11·2 per cent and those in the range from the 11th to the 40th percentage bands from the top shared 39 per cent of the total.

INCOMES BEFORE TAX

Percentile incomes before tax are shown in Table 16.1 and index numbers derived from them in Table 16.2.

TABLE 16.1

INCOMES BEFORE TAX

	1949 £	*1957* £	*1959* £	*1960* £	*1961* £	*1962* £	*1963* £
Top 1% at least	1861	2545	2839	3039	3162	3255	3364
„ 5% „	763	1198	1350	1442	1526	1592	1697
„ 10% „	565	940	1068	1151	1215	1273	1357
„ 20% „	427	773	844	909	966	1028	1067
„ 30% „	353	667	718	776	828	865	899
„ 40% „	304	577	613	653	719	742	773
„ 50% „	259	491	514	543	612	621	645
„ 70% „	122	304	315	331	374	378	400

Between 1949 and 1957, the lower the income the greater the rate of increase of income (first column, Table 16.2). Without exception, the growth rate increases as one moves from the upper to the lower income ranges. These trends, which have been studied in detail by Lydall and Paish, indicate some reduction in the inequality of incomes in the post-war years up to 1957.

TABLE 16.2

INDEX NUMBERS OF INCOMES BEFORE TAX

	1957 (1949 = 100)	*1959* (1957 = 100)	*1960*	*1961* (1959 = 100)	*1962*	*1963*
Top 1%	137	112	107	111	115	119
„ 5%	157	113	107	113	118	126
„ 10%	166	114	108	114	119	127
„ 20%	181	109	108	115	122	126
„ 30%	189	108	108	115	121	125
„ 40%	190	106	107	117	121	126
„ 50%	190	105	106	119	121	126
„ 70%	249	104	105	119	120	127

The next column, however, shows a different picture. Between 1957 and 1959 the top three percentile incomes increased most rapidly, the rates of increase of the lower ranges being progressively less, so that the lowest pre-tax incomes show the least rate of growth. This represents a movement away from equality, the higher incomes pulling away from the lower ones. Over the period 1959–63 all incomes, except the top one, increased in about the same proportion. It is clear, therefore, that since 1957 the trend towards equality in incomes before tax has ceased.

An alternative way of showing these trends, and one which brings out clearly the changes in the distribution of incomes over time, is to set out the percentages of total income held by selected groups of income recipients, counting from the richest to the poorest.

TABLE 16.3

PERCENTAGE DISTRIBUTION OF INCOMES BEFORE TAX

Group of income recipients	*1949* %	*1957* %	*1959* %	*1960* %	*1961* %	*1962* %	*1963* %
Top 1%	11·2	8·2	8·4	8·5	8·1	8·1	7·9
2%– 5%	12·6	10·9	11·5	11·4	11·1	11·1	11·2
6%–10%	9·4	9·0	9·5	9·8	9·7	9·7	9·6
11%–40%	34·9	37·6	38·4	38·5	37·6	38·6	39·0
41%–70%	19·2	23·1	22·5	22·1	23·5	22·6	22·6
Bottom 30%	12·7	11·3	9·7	9·8	10·0	9·8	9·7

The most considerable change is that which took place between 1949 and 1957. Between these two years the proportion of pre-tax

income accounted for by the top 10 per cent of incomes fell from over 33 per cent to 28 per cent, while that of the middle 11–70 per cent increased from 54 per cent to nearly 61 per cent. The share of the bottom 30 per cent fell slightly, so that, as Paish has noted, the redistribution was from the extremes to the middle ranges. The percentage shares from 1957 onwards, however, show no continuation of these trends. Apart from a reduction in the proportion going to the bottom 30 per cent, the distribution since 1957 has not significantly changed.

INCOMES AFTER TAX

Because of the stability that we have just noted in the distribution of pre-tax incomes since 1957, changes in rates of income tax and surtax may be important in affecting the distribution of incomes at the actual disposal of persons. Table 16.4 will serve to summarise the way tax rates have changed. It shows, in the form of index numbers for two specimen families of different composition, the amounts of earned income taken in tax at different levels of income for the tax years 1956–7 to 1962–3.

TABLE 16.4

INDEX NUMBERS OF AMOUNTS TAKEN IN INCOME TAX AND SURTAX

	Married couple without children					*Married couple with 3 children*			
	(*Pre-tax earned income £*)								
	600	*1000*	*2000*	*5000*	*10,000*	*1000*	*2000*	*5000*	*10,000*
1956–7	100	100	100	100	100	100	100	100	100
1957–8	100	100	100	89	92	84	97	85	89
1958–9	100	100	100	89	92	84	97	84	89
1959–60	88	91	91	83	87	73	88	79	84
1960–1	88	91	91	83	87	73	88	79	84
1961–2	88	91	91	61	66	73	88	59	64
1962–3	77	87	90	61	66	66	87	59	64

Clearly, there has been a general reduction in rates of direct taxation, but the most considerable fall has been in the rates on higher incomes since 1960–1. On an income of £1000, a married couple with three children paid in tax in 1960–1 rather less than three-quarters of what they paid in 1956–7, and two-thirds of this sum in 1962–3. If their income had been £5000, however, while in 1960–1 they would still have paid 79 per cent of their 1956–7 tax total, in 1962–3 their tax liability would have dropped sharply to 59 per cent. This greater proportionate reduction was, in fact, experienced by all those with annual incomes of about £3000 and upwards, i.e. incomes in the top ranges.

Under a progressive tax system, the effect of reductions in tax rates can be offset by the effect of inflation on money incomes. As Paish pointed out:

> During a period of rising prices, a man who succeeds in increasing his income before tax by an amount just enough to offset the rise in the cost of living, so that his *real* income before tax remains constant, is continually tending to move into higher tax brackets. With a system of progressive taxation, he will therefore pay in tax a constantly increasing proportion of a constant real income, even if rates of tax are not raised.

To show how tax changes have affected the distribution of incomes, Tables 16.5 and 16.6 repeat the foregoing analysis for percentile incomes (as defined on p. 265) and the shares of groups of income recipients for incomes after tax.

TABLE 16.5

INCOMES AFTER TAX

	1949 £	*1957* £	*1959* £	*1960* £	*1961* £	*1962* £	*1963* £
Top 1% at least	1270	1822	2073	2161	2327	2406	2545
„ 5% „	660	1038	1212	1292	1357	1427	1480
„ 10% „	520	860	965	1036	1087	1164	1205
„ 20% „	382	701	787	831	868	933	973
„ 30% „	317	593	655	711	761	794	819
„ 40% „	277	526	557	588	633	667	699
„ 50% „	250	439	478	508	540	557	576
„ 70% „	115	281	287	305	333	336	356

Incomes after tax over the period 1949–57 (except that at the 50 per cent mark) show the same trend towards equality as incomes before tax. The changes between 1957 and 1959, similarly, bring this trend to an end with higher post-tax incomes increasing more than those in the lower ranges. Between 1959 and 1963 all incomes increase by about the same amount. This pattern of change is confirmed by the distribution of incomes after tax, as given in Table 16.7.

TABLE 16.6

INDEX NUMBERS OF INCOMES AFTER TAX

	1957 (1949 = 100)	*1959* (1957 = 100)	*1960*	*1961* (1959 = 100)	*1962*	*1963*
Top 1%	144	114	104	112	116	123
„ 5%	157	117	107	112	117	122
„ 10%	165	112	107	113	121	125
„ 20%	183	112	106	110	119	124
„ 30%	187	110	109	116	121	125
„ 40%	190	106	107	114	120	126
„ 50%	176	109	106	113	117	121
„ 70%	245	102	106	116	117	124

Whether taken before tax or after tax, therefore, the distribution of incomes shows the same general trends, with the tendency towards greater equality ceasing after 1957. However, some slight difference between the two bases is shown by the results for the very high incomes. Before tax, the highest percentile income (i.e. that relating to the top 1 per cent of incomes) increased less rapidly than the others over the period 1959 to 1963: by 19 per cent, against 25–27 per cent for the rest (Table 16.2). After tax, however, the increase is the same as for the other groups (Table 16.6). Similarly, the proportion of incomes before tax taken by the top 1 per cent class fell over this period, from 8·4 to 7·9 per cent (Table 16.3), whereas on a post-tax basis the share was unchanged, at 5·2 per cent (Table 16.7). Thus, tax changes

TABLE 16.7

PERCENTAGE DISTRIBUTION OF INCOMES AFTER TAX

Group of income recipients	*1949*	*1957*	*1959*	*1960*	*1961*	*1962*	*1963*
	%	%	%	%	%	%	%
Top 1%	6·4	5·0	5·2	5·1	5·5	5·5	5·2
2%– 5%	11·3	9·9	10·6	10·5	10·5	10·7	10·5
6%–10%	9·4	9·1	9·4	9·4	9·1	9·4	9·5
11%–40%	37·0	38·5	39·8	39·8	38·9	39·2	39·5
41%–70%	21·3	24·0	23·7	23·5	24·3	23·6	23·5
Bottom 30%	14·6	13·4	11·2	11·7	11·9	11·7	11·8

underline the ending of the movement towards equality, since their effect is to improve the relative position of the top income groups – a somewhat striking result, conflicting with what might have been expected from a tax system in which rates are steeply progressive.

THE BURDEN OF TAXATION

Table 16.8 shows that between 1949 and 1957 the proportion of total tax collected from the top 5 per cent of incomes fell from 69½ to just over 54 per cent, while that raised from the middle 11 per cent–70 per cent band increased correspondingly, from 20½ to 36 per cent. This shift in the burden of taxation reflects the combined effects of three factors: changes in the tax rates, which

TABLE 16.8

PERCENTAGES OF TOTAL INCOME TAX AND SURTAX RAISED FROM SPECIFIED GROUPS OF INCOME RECIPIENTS

Group of income recipients	*1949*	*1957*	*1959*	*1960*	*1961*	*1962*	*1963*
Top 1%	46·0	35·3	34·5	34·4	30·0	28·4	28·0
2%– 5%	23·6	18·9	20·1	19·7	19·9	18·5	19·3
6%–10%	8·6	9·0	10·2	10·0	10·0	10·8	11·0
11%–40%	16·2	24·9	24·8	25·4	27·6	29·7	29·8
41%–70%	4·4	11·2	9·7	9·7	11·4	11·8	11·3
Bottom 30%	1·3	0·7	0·6	0·8	1·0	1·0	0·6

were reduced during these years; the inflation of money incomes – so that equivalent real incomes are taxed at higher proportionate rates; and the redistribution of pre-tax income away from the highest income groups taxed at the highest rates, a development which Paish has called the 'loss of taxable capacity'. This loss of taxable capacity (shown by the figures in Table 16.3) was the main cause of the shift. Although tax rates on the highest incomes fell proportionately less than those on lower incomes, the percentages of pre-tax income taken in direct taxes (shown in Table 16.9) fell for the top 5 per cent of incomes but rose or remained constant for the 6 per cent–70 per cent band.

Between 1957 and 1963 the tax burden continued to be lightened on the highest incomes. The portion of total tax collected from the top 5 per cent of incomes fell from 54 to 47 per cent, but that from the 6 to 10 per cent band rose from 9 to 11 per cent and that from the 11 to 70 per cent band increased from 36 to 41 per cent

(Table 16.8), the major part of these shifts occurring after 1960. How did this happen?

TABLE 16.9
PERCENTAGES OF PRE-TAX INCOME OF SPECIFIED GROUPS OF INCOME RECIPIENTS TAKEN IN DIRECT TAXES

Group of income recipients	*1949*	*1957*	*1959*	*1960*	*1961*	*1962*	*1963*
Top 1%	48·8	45·5	43·1	46·1	41·5	39·7	39·1
2%– 5%	22·4	18·4	18·3	19·7	20·0	19·9	19·0
6%–10%	10·9	10·7	11·3	11·6	11·5	12·6	12·5
11%–40%	5·5	7·1	6·8	7·5	8·2	8·8	8·4
41%–70%	2·7	5·2	4·5	5·1	5·4	5·9	5·5
Bottom 30%	1·2	0·7	0·7	0·9	1·1	1·1	0·6
All incomes	11·9	10·6	10·5	11·4	11·2	11·4	11·0

During these years, as we have seen, the distribution of income before tax was virtually unchanged. There was, therefore, no further significant loss of taxable capacity. The shift in the tax burden must, then, have been the result of the greater proportionate reduction after 1960 in tax rates on the highest incomes (Table 16.4). Indeed, as Table 16.9 shows, it was only for the top 1 per cent that there was a fall in the percentage of income taken in direct taxes over the period 1957 and 1963, a fall proportionately greater than that between 1949 and 1957. For all other ranges of income the reduction in tax rates has been offset by the inflation of money incomes, with a corresponding rise in the relative amounts taken by the tax collector.

Thus, although the burden of direct taxation has shifted over the whole period 1949 to 1963, the explanation of the shift has changed. Before 1957 it was due to loss of taxable capacity in the highest income groups; since then it has been due to changes in tax rates.

REASONS FOR CHECK TO GREATER EQUALITY

The trend towards the reduction of inequality in the distribution of personal income seems to have come to an end by 1957. Indeed if, as is sometimes suggested, certain 'tax avoidance' incomes and other claims on wealth outside personal income have increased over the last decade and are concentrated more among higher-income recipients, it is possible that the distribution of incomes on some wider definition may have moved towards greater inequality. This, however, is conjectural. In the context of personal income as defined in the national income accounts, a comparison of Tables 16.3 and 16.7 shows that the significant change has been the end of the trend towards equality in incomes before tax. Why has this occurred?

The reasons for the continued movement towards equality after the war were rising prices, continuously high demand and the virtual elimination of unemployment, all of which led to earned income rising faster than other forms of personal income. Lydall's view was that, so long as these trends were unchecked, the tendency towards equality would continue. There was certainly no failure of demand after 1957 – gross domestic product grew at a marginally faster rate over 1957–63 than over the years 1949–57 – and employment was no less full. What, then, has happened? It is true that consumers' prices increased at a faster average rate in the years 1949 to 1957 than in the later period 1957 to 1963: at 3·9 per cent, against 2·0 per cent, a year. This may have been a contributing factor to the end of the movement towards equality, but is unlikely to be the whole explanation. More important has been the change in the rates of growth of different forms of personal income.

TABLE 16.10
GROWTH OF FORMS OF PERSONAL INCOME

	1938–49 (1938 = 100)	*1949–57* (1949 = 100)	*1957–63* (1957 = 100)
Wages	231	173	129
Salaries	237	190	157
Total employment income	240	179	140
Professional persons	189	129	136
Farmers	509	129	118
Sole traders	179	129	125
Total self-employment income	216	129	125
Rent, dividends, interest	104	142	172
Transfers	271	170	179

The first two columns, covering the periods during which reductions in inequality of personal income have been demon-

strated by Paish and Lydall, show that the most rapidly growing sector of personal income was income from employment, growing faster than total self-employment income and incomes from rent, dividends and interest.

Since 1957 the pattern has changed. Growth of employment income has slowed down relatively to that of self-employment income and, within employment income, the rate of growth of wages has slowed down relatively to that of salaries. In self-employment income, that of professional persons has expanded most rapidly, faster, in fact, than wages. The most striking change, however, has been the accelerating rate of growth of rent, dividends and interest after the war, emerging as the most rapidly growing sector of personal income since 1957.

Some light on the relative growth rates of total wages and salaries is shed by the figures below, derived from the national income accounts, showing numbers of wage- and salary-earners, and average wages and salaries in manufacturing industry.

TABLE 16.11

WAGES AND SALARIES IN MANUFACTURING INDUSTRY

	1949	*1955*	*1957*	*1963*
Wage earners ('000)	5870	6340	6290	6150
Salary-earners ('000)	1170	1500	1600	1890
Average wage (£)	290	444	503	659
Average salary (£)	489	661	754	1020

Between 1949 and 1957 the average wage in manufacturing increased by 73 per cent and average salary by only 54 per cent. Up to 1955, moreover, the number of wage-earners increased more than the number of salary-earners. Since 1957 average salaries have caught up, and increased at much the same rate as average wages: between 1957 and 1963 the increases in manufacturing industry were 35 per cent for salaries and 31 per cent for wages, rates of growth confirmed by the Ministry of Labour's indexes of salaries and average earnings (which include industries and services outside manufacturing). The number of wage-earners, however, has fallen since 1955, whereas the number of salary-earners has continued to rise, indeed by nearly one-fifth since 1957. Thus, with the average salary increasing as fast as the average wage, the relatively greater rate of growth of *total* salaries than of wages since 1957 is due to the continued increase in the number of salary-earners. Table 16.11 shows also that the average salary is about 50 per cent greater than the average wage but, of course, salaries reach up into much higher income ranges than wages.

It is not possible to obtain estimates of numbers or average incomes of those classified in the national income accounts as professional persons, but since these include doctors, dentists, lawyers and architects, average professional incomes are certainly very much higher than wages. Rent, dividends and interest, like salaries and professional incomes, are usually associated with higher incomes and it is, therefore, the more rapid growth of these three forms of personal income, particularly rent, dividends and interest, compared with that of wages, that has ended the trend towards equality in income.

CONCLUSIONS

Dr Lydall, writing of the years up to 1957, saw the trend to greater equality, though unlikely to continue as rapidly in the future as in the immediate past, as a long-term development which might be stopped or reversed only by the onset of a major slump. Professor Paish was more doubtful:

> It seems likely, however, that the process of redistribution would tend to slow down even if prices continued to rise. . . . Most of all it could be because investors and others . . . will increasingly seek out means for ensuring that their money incomes rise at least part of the way with them.

The more rapid growth of the incomes of professional people and of rent, dividends and interest since 1957 supports Paish's view. A particular instance of the latter is seen in the increased buying of equities by private individuals, either directly or through unit trusts specialising in high-income yields adding to income or in capital growth adding to wealth (which is more unequally distributed than income) and making for future capital gains. How far these tendencies will be offset by the capital gains tax or by changes in rates of income tax which, under a Labour Government, may reverse recent trends and become more steeply progressive, it is impossible to say. The corporation tax introduced in April 1965 is intended to give a strong inducement to companies

to plough back more of their profits for expansion. It is possible, therefore, that the proportion of company income distributed as dividends – which has increased over the past decade – may decline. Nevertheless, the aggregate amount distributed as dividends can be expected to increase as total profits increase, so that personal income from dividends will continue to grow even if, possibly, less rapidly than in the past.

At the other end of the scale, the increases in pensions introduced in 1965 will raise some incomes in the lower ranges, but it is unlikely that this will bring about any fundamental change in the distribution of incomes, because relatively small amounts are involved and increases in pensions tend to lag behind increases in other incomes. Certainly, the present deflationary policies, the incomes freeze and the slowing down of the growth of the national product, following the stability of the income distribution since 1957, must make it doubtful whether the trend in the distribution of incomes for the immediate future is set in the direction of greater equality.

Notes

1. 'The Real Incidence of Personal Taxation', *Lloyds Bank Review* (January 1957).
2. 'The Long-Term Trend in the Size Distribution of Income', *Journal of the Royal Statistical Society*, series A, vol 122 pt 1 (1959).
3. *Income Distribution and Social Change* (Allen & Unwin, 1962).
4. This method of using 'percentiles' was that employed by Lydall. Alternatively, using Paish's method, the comparisons could be made in terms of the shares of specified *numbers* of incomes. In view of the gradual increase in the number of incomes over the years, the percentile-income method is preferable. All the numerical results given in the present article have been obtained by logarithmic interpolation into the published distributions.

This piece is a short extract from 'Religion and Society among the Coorgs of South India' by the eminent Indian sociologist Professor M. N. Srinivas. The book was published in 1952 – so one must remember that it is an account of Indian caste about one generation ago.

A 'caste' is usually marked by endogamy (members may not marry outside the caste); membership is hereditary, and the impact of a caste upon many aspects of its members' lives is profound. Caste involves an especially rigid form of stratification, with the implication that each man must accept the place in society to which he is born; caste has strong religious overtones, it is linked to occupational specialisation and pervades all aspects of a man's style of life. Castes are socially segregated from each other – and in this process notions of ritual pollution are important. As an extreme example, a high-caste Hindu might throw away his food, should the shadow of an 'Untouchable' fall upon it.

But as Professor Srinivas showed in his study, the caste system does not have the geological kind of rigidity which appears at first sight. The number of castes and sub-castes is vast, and the exact ranking of castes varies somewhat according to locality. Even in this – the most pervasive, comprehensive and complex system of social stratification – there is some room for change. Castes do sometimes manage to improve their position.

Although economic and political changes in India since 1950 have modified the picture presented here, something as fundamental and deeply entrenched as the caste system was in India cannot be removed within a single generation. Indeed there is some evidence that the caste system has actually been strengthened in recent years, at least in rural areas.

THE CASTE SYSTEM IN INDIA

M. N. Srinivas

Caste is an institution of great complexity. It has its roots deep in history, and even today it governs the lives of 300 million Hindus in several important respects. It is popularly understood as the division of society into a fivefold hierarchy with the Brahmins at the head, followed in order by Kshatriyas, Vaishyas or traders, Shudras or servants and labourers and, lastly, the Untouchables. The first three castes are called 'twice-born' (*dvija*) as they alone are entitled to undergo the ceremony of *upanayana* which constitutes spiritual rebirth. Only the twice-born castes are entitled to study the Vedas and do the performance of Vedic ritual on certain occasions. Caste in the above sense is referred to as *varṇa* and has an All-India application.

The idea of caste as the fivefold division of society represents a gross oversimplification of facts. The real unit of the caste system is not one of the five *varṇas* but *jāti*, which is a very small endogamous group practising a traditional occupation and enjoying a certain amount of cultural, ritual and juridical autonomy. Every *jāti*, or the members of a *jāti* in a particular village or a group of neighbouring villages, constitutes a caste court which punishes caste offences.

There are innumerable *jātis*. Professor Ghurye calculates that there are 2000 sub-castes (*jātis*) in each linguistic area.[1] This should give some idea of the total number of endogamous sub-castes in India as a whole. The importance of the *varṇa*-system consists in that it furnishes an All-India frame into which the myriad *jātis* in any single linguistic area can be fitted. It systematises the chaos of *jātis* and enables the sub-castes of one region to be comprehended by people in another area by reference to a common scale. Further, the *varṇa*-system represents a scale of values, and *jātis* occupying the lower rungs have throughout tried to raise their status by taking over the customs and ritual of the top *jātis*. This has helped the spread of a uniform culture throughout Hindu society.

The attempt to fit the *jātis* of any region into the fivefold hierarchy is a very difficult affair. It is possible everywhere to say who are the Brahmins and Untouchables, but there is great confusion in the middle regions. Confining our remarks to South India for the moment, we find that the claims of local *jātis* to be Kshatriyas and Vaishyas are frequently questioned by others. For instance, it is well known that a ruling house claiming to be Kshatriyas were originally potters. Similarly, the claim of a local

trading *jāti* to be Vaishya, one of the three twice-born *varṇas* of the Vedas, is hotly contested by the other castes in the area. It is not uncommon to find a *jāti* included under the Shudra *varṇa* claiming to be higher than the local *jātis* claiming to be Kshatriyas and Vaishyas. Disputes as to relative status are an essential feature of the caste system. This is especially so in the numerous *jātis* belonging to the fourth *varṇa* of Shudra. In south India the term 'Shudra' includes the vast majority of non-Brahmin *jātis* and even some reformist sects.

A man is *born* into a sub-caste (*jāti*) and this is the only way of acquiring membership. According to the traditional view, however, birth is not an accident. Certain Hindu theological notions like *karma* and *dharma* have contributed very greatly to the strengthening of the idea of hierarchy which is inherent in the caste system. The idea of *karma* teaches a Hindu that he is born in a particular sub-caste because he deserves to be born there. The actions he performed in a previous incarnation deserved such a reward or punishment, as the case might be. If he had performed better actions in his previous incarnation he would have been born in a higher caste. Thus the caste hierarchy comes to be an index of the state of an individual's soul. It also represents certain milestones on the soul's journey to God.

Thus the idea of deserts is associated with birth in a particular caste. A man is born in a high caste because of the good actions performed by him in his previous life, and another is born into a low caste because of bad actions performed in his previous life.

The other important concept is *dharma*, which has many meanings, one of which is 'that which is right or moral'. The existing moral code is identified with *dharma*. A man who accepts the caste system and the rules of his particular sub-caste is living according to *dharma*, while a man who questions them is violating *dharma*. Living according to *dharma* is rewarded, while violation of *dharma* is punished, both here and hereafter. If he observes the rules of *dharma*, he will be born in his next incarnation in a high caste, rich, whole and well endowed. If he does not observe them he will be born in a low caste, poor, deformed and ill endowed. Worldly position and success indicate the kind of life a man led in his previous incarnation.

One may also reap the reward of one's actions very soon after their performance. For purposes of such reward and punishment, a person is identified with his joint family. A man may become blind because his father, the head of the joint family, made money in the black market during the war.

The concept of pollution governs relations between different castes. This concept is absolutely fundamental to the caste system, and along with the concepts of *karma* and *dharma* it contributes to make caste the unique institution it is. Every type of inter-caste relation is governed by the concept of pollution. Contact of any kind, touching, dining, sex and other relations between castes which are structurally distant results in the higher of the two castes being polluted. Ordinarily, contact between members of the same caste, or between members of castes which are structurally very near each other, does not result in pollution. Where contact does result in pollution, however, the polluted member of the higher caste has to undergo a purificatory rite in order to be restored to normal ritual status. Such a purificatory rite is fairly simple where the structural distance between the castes is not very great and the type of contact is not serious. Sometimes, as when a Brahmin eats food cooked by an Untouchable, the resultant pollution is so great that he or she has to be excommunicated. Normally, in every caste, women observe the pollution rules much more strictly than men.

Contact is culturally defined. Touch is contact in all cases. Frequently the maintenance of a minimum distance between castes is insisted upon. This matter has received systematisation in Kerala, where elaborate rules have been laid down requiring the minimum distance that should prevail between the various castes.

> Aiyappan, in 1937, gives a scale of distance pollution for several castes: a Nayar must keep 7 ft from a Nambudri Brahman, an Iravan (Ilavan, Izhuvan, Tiyan) must keep 32, a Cheruman, 64 and a Nayadi 74 to 124. The respective distances between these lower castes are calculated by a simple process of subtraction: the Iravan must keep 25 feet from the Nayar and Cheruman 32 feet from the Iravan.[2]

The ban on contact between castes, and the solidarity of a sub-caste, express themselves in the spatial segregation of castes in a village. The Untouchables live everywhere at some distance from the others, while each of the other sub-castes occupies a street or a quarter of the village or town.

Normally, a man may accept cooked food and water from a member of the same or equal or superior caste. Food cooked by a member of a lower caste may not be eaten because such food defiles a man belonging to a higher caste. Mutual acceptability of cooked food denotes equality between the castes concerned, while the movement of food in one direction only indicates that the acceptor is inferior to the giver.

People living in rural areas, and the orthodox, exhibit a preoccupation with the matter of acceptance or rejection of cooked food from different castes. The kind of food, the question whether it is cooked in butter or water, the caste of the person cooking it, and the place (whether temple or home) where the food is cooked, all go to determine its acceptability or otherwise. A tendency to define and systematise is obvious everywhere.

There is a general correlation between diet and status. Brahmins are usually vegetarians, abstaining even from eggs. The Shudra castes eat eggs and meat. Pork is inferior to mutton, and usually pork-eaters to mutton-eaters. Beef-eaters are the lowest of all. The consumption of toddy, too, begins at some point low down in the caste hierarchy.

Facts have been oversimplified in order to make some general statements. For instance, while Brahmins are usually vegetarians, the Saraswat Brahmins of the west coast, and some Brahmins of Bengal, eat fish, while the Kashmiri Brahmins eat meat. Again, castes claiming to be Kshatriyas generally eat mutton. The Lingayats of south India, who are non-Brahmins, are vegetarians and teetotallers. It must be mentioned here that from the earliest times vegetarianism has been held up as an ethical ideal, it being considered wrong for an individual to kill other sentient creatures in order to keep himself alive. Vegetarianism goes hand in hand with teetotalism, the drinking of alcoholic liquors being forbidden to the high castes.

The acceptance of cooked food by a high-caste man from an Untouchable constituted as serious an offence as having a liaison with an Untouchable, and in the old days both offences were punished, by the caste court concerned, with excommunication from caste. It is not only the high castes who are particular about the caste of the person from whom they accept food and water, but the Untouchables too are very particular. In Mysore, Untouchables do not accept food and water from smiths, and from Marka Brahmins. They believe such acceptance would defile them.

Each caste is traditionally associated with a separate occupation. Some of the earlier students of caste were so impressed with this feature of caste that they ascribed the origin of caste to the systematisation of occupational differentiation. In rural India the bulk of the castes continue to practise their respective traditional occupations, though agriculture is common to all castes from the Brahmin to the Untouchable.

Some occupations are considered defiling because of the contact with some defiling object or other necessary to their practice. Swine-herding is defiling because swine defile. Leather defiles, and consequently the making and repairing of shoes is an occupation of the Untouchables. Certain other occupations, while not being defiling, are so strongly associated with low castes that no high-caste man will take to them. The tapping of toddy would be a case in point. Finally, any occupation however remotely implying the destruction of sentient life in any form would be prohibited to the high castes. Thus not only butchery, the catching of fish and game, and the raising of sheep and fowls would be prohibited, but even the selling of dried fish and eggs would be regarded as an improper occupation for a high-caste man, especially a Brahmin. But nowadays, under the influence of industrialisation, the old association of caste with occupation is beginning to break down.

The members of a sub-caste tend to regard their traditional occupation as the natural one for them. Taking up any other occupation is regarded as improper. There is a pride in the skill required for the practice of the traditional occupation, and this skill is a secret which is not easily divulged to members of other castes. In fact, formerly castes tended to be guilds, and the members of castes regarded their traditional occupation as their monopoly. A violation of this monopoly would result in a fight. The matter could also be taken before the village court and before the king of the region.

Things associated with high caste, their houses, clothes, customs, manners and ritual, tend to become symbols of superior status. Consequently there is a ban on the lower castes taking them over and, formerly, such a ban was sanctioned by the political authority at the top. In Malabar, until 1865, only the

Brahmins were allowed to clothe their bodies above the waist, even the women of the lower castes being prohibited from doing so. In many parts of India the lowest castes were not allowed to have tiled roofs for their houses. Nor were they allowed to build two-storied houses.

A consequence of the extreme stratification implied in the caste system is the tendency of each sub-caste, or each level of sub-castes, to live in a separate social world. The members of a sub-caste inhabit the same quarter of the village or town and frequently are all related to each other by agnatic or affinal links. They share a common culture and ritual idiom. They observe common restrictions regarding food and drink, and have certain caste festivals and rites not shared with others. They practise a common, traditional occupation, the secrets of which they do not share with others. They have caste courts and assemblies where elders of the sub-caste belonging to different villages assemble and decide matters of common concern. The members of a sub-caste share certain common values and are actively aware of this fact when they come into contact with other castes. The autonomy of a sub-caste does not, however, mean that it can live independently of other sub-castes. An Indian village usually consists of a few sub-castes which are mutually dependent and also possess certain interests in common.

The sub-caste or *jāti* is the unit of endogamy. Occasionally we find a *jāti* split up into several groups, each of which is endogamous. The rule about endogamy is still strictly observed, though a contemporary tendency among the educated members of the higher sub-castes is to marry into another sub-caste of the same level and sharing the same culture. Thus a member of the Mandyam division of the Sri Vaishnava Brahmins might nowadays marry a member of the Hebbar division of the same larger group, though formerly they would not have done so.

Hypergamous unions occasionally occur between castes. By this a man belonging to the higher caste takes a girl from the lower caste. It is never the other way about. Hypergamous marriages are common in Malabar and Bengal. They occur in other parts of India too.

The caste system is far from a rigid system in which the position of each component caste is fixed for all time. Movement has always been possible, and especially so in the middle regions of the hierarchy. A low caste was able, in a generation or two, to rise to a higher position in the hierarchy by adopting vegetarianism and teetotalism, and by Sanskritising its ritual and pantheon. In short, it took over, as far as possible, the customs, rites and beliefs of the Brahmins, and the adoption of the Brahminic way of life by a low caste seems to have been frequent, though theoretically forbidden. This process has been called 'Sanskritisation' in preference to 'Brahminisation', as certain Vedic rites are confined to Brahmins and the two other 'twice-born' castes.

The tendency of the lower castes to imitate the higher has been a powerful factor in the spread of Sanskritic ritual and customs, and in the achievement of a certain amount of cultural uniformity not only throughout the caste scale, but over the entire length and breadth of India.

Another point which needs specially to be stressed is that the idea of hierarchy embodied in the caste system has been periodically questioned in India. Buddhism and Jainism questioned it quite early in the recorded history of caste. Lingayatism challenged it in the south in the twelfth century A.D. These attempts more or less met with failure, and occasionally the reformist sect either became a caste or reproduced within itself a caste system.

Many of the sanctions supporting the caste system tended to disappear under British rule. The British withdrew the explicit as well as implicit support which caste usually enjoyed under an Indian monarch. For instance, people who had been punished by caste courts could sue the caste elders responsible for defamation in courts established by the British.

Formerly, an offender against the code of the caste was punished with either a fine or, in very serious cases, excommunication from caste. Between themselves, caste councils and village councils completely controlled the conduct of an individual. Caste elders and village elders supported each other's authority.

Caste guarantees autonomy to a community, and at the same time it brings that community into relation with numerous other communities all going to form a hierarchy. The importance of such an institution is obvious in a vast country like India which has been the meeting-place of many different cultures in the past and which has always had considerable regional diversity. While the autonomy of a sub-caste was preserved it was also brought into relation with others, and the hierarchy was also a scale of generally

agreed values. Every caste tended to imitate the customs and ritual of the topmost caste, and this was responsible for the spread of Sanskritisation. When this process is viewed on a continental scale and over a period of at least 2500 years, it is easy to see how Sanskritic ideas and beliefs penetrated the remotest hill tribes in such a manner as not to do violence to their traditional beliefs. Caste enabled Hinduism to proselytise without the aid of a church.

Caste, which was so successful in absorbing autonomous groups everywhere, also provided the pattern for relations with non-Hindu groups. Christians and Muslims were regarded as castes, too, and they accepted such a status. Even revolutionary movements which had aimed at the overthrow of the caste system ended by either becoming castes themselves or reproduced the caste system within themselves. The main body of Hindus regarded these sects as castes and not as sects. Thus the caste system effectively neutralised all attempts to change it.

Notes

1. G. S. Ghurye, *Caste and Race in India* (London, 1932).
2. A. Aiyappan, *The Anthropology of the Nayadis* (1937).

The social stratification of society has received far more attention from sociologists than from psychologists. This is not because psychological problems are lacking in this area; more probably the apparent dearth of interest has been due to the fact that it is difficult to bring the psychological phenomena of social class into the laboratory for closer examination.

Nevertheless, over the years Professor Pear and a few others have accumulated a great deal of relevant source material that will doubtless provide a springboard for the next generation of social psychologists. This is the first step in the application of scientific method – systematic observation. The next step is to organise these findings within a theoretical framework, and to some extent this is already happening. From it follows the setting-up of testable hypotheses.

PSYCHOLOGICAL ASPECTS OF ENGLISH SOCIAL STRATIFICATION

T. H. Pear

THE stratification of our society, though obvious and striking, is hardly mentioned in English psychological textbooks, except those by J. M. Blackburn, R. H. Thouless and W. J. H. Sprott.[1] English sociologists Margaret and G. D. H. Cole, Morris Ginsberg, T. H. Marshall, H. A. Mess, R. H. Tawney and Barbara Wootton have discussed this subject, and recently R. Lewis and A. Maude have added valuable contributions to this collection.

Our surprise at this neglect by psychologists may be lessened by the realisation that this tepidity of interest itself offers psychological problems. Social forces make it difficult for a professional thinker to regard his own position in society objectively, and fully to comprehend its meaning in a social pattern.

In most societies, stratification exists, yet upon the English it exerts a powerful, subtle influence. In younger nations matters are less complex. In the U.S.A., as Professor D. W. Brogan has pointed out, social stratification is important, but its delimitations and the forces producing these differ considerably from those known in England. He writes:

> It is only an apparent contradiction in terms to assert that the fundamental democratic and egalitarian character of American life is demonstrated by the ingenuity and persistence shown in inventing marks of differences and symbols of superiority. In a truly class-conscious and caste-dominated society, the marks of differences are universally recognised even if resented. In America they must be stressed, or they might easily be forgotten, and they must be added to, as the old standards of distinction cease to serve their purpose. Apart from the simple economic criterion of conspicuous display, there are no generally accepted marks of social differences in America. And modern salesmanship makes clothes, cars and personal adornment far more alike than was possible in the old days of belated styles and the Model T Ford. It is worth noting that the main stress of American class distinction is put on 'exclusiveness'. In a society without formal public recognition of differences in rank, with a poor and diminishing stock of natural reverence for hereditary eminence and with a constant rise to the top of the economic system of new men amply provided with the only substitute for hereditary eminence, wealth, it becomes extremely difficult to make 'society' anything but the spare-time activities of the rich. It is characteristic that it is in cities whose days of economic advance are over, in Boston, Philadelphia, Charleston, that it has proved easiest to keep out the newcomers.

In some parts of England the relation between social and economic status is complex. A poor country vicar may often visit rich neighbours, but he does not 'mix with', as distinct from visiting parochially, many parishioners whose incomes may equal or surpass his own. In the intellectual professions, especially in large towns with literary and musical coteries, within certain limits social status may be variable almost at will. Yet nowadays wealth is an important distinguishing factor of the English upper strata, since few impoverished aristocrats can live entirely on credit.

In this country, economic status is easy to guess at. Though to indicate it overtly, while referring to anyone by name, is 'not done', euphemisms – 'comfortably off', 'in reduced circumstances' – hint with fair accuracy at financial position. Nevertheless, widespread severe taboos are imposed upon detailed consideration of social status (especially of persons living in any but the 'top' and 'bottom' layers of society), and of speaking and clothing, the chief signs of class. These prohibitions are overt in upper and middle-class society, where 'gentlemen' and 'ladies' are seldom referred to as such, except perhaps by old people. Moreover, at a less conscious level, any attempt to describe or record the outstanding characteristics of gentlemen and ladies would be resented and resisted. As a Continental sociologist put it, there has been little field-work among the English upper classes. Similar taboos probably exist in the minds of some social philosophers who consider themselves immune to such foolishness.

This unanalysed acceptance of social stratum ('knowing one's place') partly explains the sociologists' and psychologists' neglect to consider social stratification, clothing and speaking. No British psychologist, except Dr R. H. Thouless, who made a promising start with social stratification, and Professors J. M. Blackburn and W. J. H. Sprott, who devote chapters to it, have thought these subjects worthy of detailed treatment in a textbook. English social psychology is, on the whole, an account by middle-class writers of middle-class behaviour. First-hand studies have been made in England by 'foreign' observers from Poland (Dr F. Zweig), the U.S.A. (E. W. Bakke), and Austria (Dr Marie Jahoda).

Psychologists ought to regard social stratification objectively, as a geologist sees layers in a cliff-face, unprejudiced about the desirability of their relative positions, even of their existence. The analogy is not far-fetched. An artist's painting of a stratified cliff might omit, widen or narrow some layers, and transpose others, as some describers of English society are tempted to do. In society, too, the boundaries between layers may be blurred, oblique or curved, while until recently the B.B.C.'s Third Programme produced a fine example of vertical stratification, since in some towns within a few miles of special aerials, reception is reasonably good on a cheap set, while in many regions such a result cannot be bought.

Dr Thouless describes significant aspects of social stratification. Among them are the social obligations of a group, summed up in its code of manners. Salutations at meeting and parting are a good example of differences in procedure in various strata of our own society. A code of manners facilitates social intercourse and the adoption of the attitude of primitive comradeship within the group. It also often increases (and this is important for our present subject) the separateness of groups from others. Dr Thouless followed a current usage speaking of 'superior' and 'inferior' classes, but thoughtfully added:

> The terms 'superior' and 'inferior' social classes will be used frequently. It is necessary to make clear that these terms are used in a severely technical sense to mean only superiority and inferiority in the scale of social stratification. Stratification is a fact of social organisation which we can recognise by various marks, perhaps most easily by noticing that the member of a superior social class gives orders to one of inferior social class while the other obeys them, and that there is a social obligation on the member of the inferior social class to use gestures or methods of address indicating respect which the other is not under the same obligation to reciprocate. Many of the differences are external signs of a relatively intangible but very real difference in social prestige attached to different classes. The use of the term 'superior' in this connection does not imply that the member of a socially superior class is superior in intelligence, social usefulness, or in any other respect except that of occupying a higher position in a stratified social system.

Manners, as he points out, may be reciprocal or non-reciprocal:

> They may also serve to mark the relationship of superior and inferior in different social strata. Between equals the social obligations of manners are reciprocal – an obligation of behaviour or speech of A towards B is equally an obligation of B towards A. Other social obligations, however, are non-reciprocal. A is required to adopt a certain form of behaviour or speech towards B who is socially superior, although this obligation does not exist or exists in a different form from B to A.

Few writers have subsequently used the terms 'superior' and 'inferior', possibly because to many readers, these words would

have just that unpleasant aura of which Dr Thouless tried to rid them. We shall not employ these adjectives, though agreeing that they are logically suitable.

Perhaps one of the most important and least often mentioned aspects of English social stratification is the inferiority of the seller in the buyer–seller relationship. This principle determines the line drawn between people 'in trade' and those in the professional classes – lawyers, accountants, etc.

A sharp stratum-line between wholesale and retail trade is usually observable. The reason why, in peace-time with free competition, the retail seller may feel inferior, in both the narrow and the wide senses of this term, seems to be clearly expressed in the maxims 'Caveat emptor!', 'Let the buyer beware!' and 'The customer is always right'. However independent in manner a retailer wishes to be, he can seldom risk losing trade. Reluctant recognition of post-war social change was perhaps expressed in a notice on a laundry-van, 'We are now prepared to accept applications from new customers', and in overt disapproval, by some suppliers of coal and milk, of the Government's permission of a change of retailer. In the U.S.A. it was an English writer who noted that in a store a customer's opening speech is apt to begin with 'Gimme!' and not to end with 'please!' A contrast in manners can sometimes be seen in an English pharmacist's if specially dispensed medicines are sold at the same counter as ready-made drugs and toilet requisites. Those who remember the manners, and charges, at a pre-war German or a present-day Swiss *Apotheke* will understand why.

This discussion may annoy some readers, but may demonstrate that it concerns a matter considered important by some purchasers. At times, the retailer must be very nearly all things to all men; the wholesaler's politeness can be more impersonal. American journalists who occasionally write 'clients' for 'customers', and vice versa, illustrate a type of humour which does not meet with universal approbation in, say, the Inner Temple or, perhaps, Bond Street.

Let us try to summarise current conflicting opinions about our present and future social stratification. Our aim is to show not only that the practical issues are of prime importance for our future welfare, for that is obvious to thoughtful people, but that where these questions are not economic they are psychological. We will then discuss psychological aspects of the general problem, hoping that it may help those who wish, while drastically changing our social system, to retain its useful features.

Here social psychology verges upon economics: many differences between English social classes are economically conditioned. Not all people in the lowest and lower-middle ranges of the economic scale, however, if they became wealthy, would choose to adopt all the customs and conventions common among the rich. Some of these usages would seem foolish, insincere, irksome or wicked to the newly-rich, who would have brought along their own prejudices, many of them also irrational.

The reader may reasonably assume that at the date of writing English social stratification is not what it was in 1939, and that it changed considerably when peace came. This has been kept in mind, but as there are few data about social changes during the war, little time can be spent upon them here. It is, however, important to record that, as in Disraeli's time, there are still two Englands. Today they may perhaps be symbolised by the one of Miss Alice Duer Miller's *The White Cliffs* and the other, explicitly distinguished from it by Mr J. B. Priestley, in *Out of the People* and *English Journey*. He contrasts, as pictures and as social patterns, the South and the North, the counties with leisured beauty and those with the crowded, noisy, dirty, ugly but useful towns. Though people often try to silence any mention of the corresponding social differences, or even pretend that they do not exist, they are immediately observed and commented upon by visitors from abroad, especially from the Dominions or the U.S.A. Acute awareness of these differences was still evident in the questions asked before Dunkirk of candidates for commissions in some of our fighting services.

From many directions we heard that after the war there would be great changes in the English social classes. In one sense this is a platitude. All modern wars cause social change; this is what they are fought for. In another sense, the statement is more interesting. It usually implies that the differences of social stratification will be fewer, and the lines of demarcation less clear. This prophecy is received by many with a sigh for the pre-war state, which they regarded as historically interesting, beautiful, romantic, desirable. Even if they personally played no actual part in the grand world, they saw it as a fairy-land, an escape-country or a goal, if not for themselves, for their children or grandchildren. At

the beginning of World War II some urged that it should be fought for the 'English way of life', describing it in a manner which might be recognised by their friends, if not by the majority of the English. Since then, many have pointed out that no national way of life ever completely survives a great war; that when peace is declared, relatively few could wish our social structure to be restored, cracks, moss, mildew and all, and the advocates of the *status quo* have been almost silent. We may, however, distinguish between writers who feel certain that great social changes, not named precisely, are inevitable, and those who suggest definite innovations.

Mr Priestley in 1941 opposed Walt Whitman's concept of 'the people' to that of 'masters' and 'masses'. We are not, or ought not to be, he wrote, conducting the war as members of different classes, but as people. 'The people can . . . be praised for their conduct so far in the war. . . . On the whole, the classes as such have behaved rather badly, being more concerned about their own particular rights and privileges than about the extent and force of our war effort. On the other hand, it is the people and not the classes who have so magnificently manned our air-raid services.' The people, he insists,

> are not the masses. These are two different conceptions. Indeed, they are opposed. One rejects the other. When I say to myself 'the people', I have a confused but lively vision of a hundred faces and a hundred voices, as if a picture by old Breughel had suddenly come to life. In short, I think of persons. But when I say to myself 'the masses', I see at once a grey, featureless horde, and hear nothing but a muttering and murmuring. I do not think of persons. The masses are not real human beings. They have been de-humanised. The supreme mystery does not flower in them as it does in us, who always feel ourselves separated from them. They are rough bundles of instincts and appetites. They are units of man-power. After using their muscles and collecting their pennies, you can consider and gratify their immediate needs; you cannot love them. They are not real persons.
>
> We have been searching for years now for that division of the modern world that cuts the deepest. Once it was socialism against capitalism; then it was fascism against communism. Lately it has been totalitarianism against democracy. A welcome change from these is the division between those who think in terms of the masses and those who think in terms of the people. I believe it cuts very deep indeed.

Some of these arguments may seem irrelevant to one who, while actively rejecting the concept of masters and masses, firmly believes (both pragmatically and theoretically) in the 'classes'. He would state, as postulates not axioms, that the class-division still exists here, that on the whole it has worked fairly well and that it is a social psychologist's business, by noting, describing and explaining the fact, to help the classes to co-operate without friction and jealousy. He may regard the impulse to rise socially as justifiable and useful.

Social stratification, however, is not the result of forces as inevitable as those which divide into neat layers a mixture of liquids of different density. It is less easy to map in a large industrial town than in a village. Mr Priestley asserts that even before the war the division into classes had lost nearly all political and economic (though not purely social) reality, that nowadays a good deal of it is merely an ornamental cover, bearing much the same relation to the basic realities of modern life that fancy dress does to ordinary clothes. As evidence, he cites the selling of titles in exchange for contributions to the funds of a political party, and Mrs Barbara Wootton's investigation (quoted on pp. 284–5) which in 1959 concluded that today the fundamental difference between classes was one of wealth. He claims that the real aristocratic system vanished years ago, and what we have had since is a plutocratic system pretending to be aristocratic. It allows the strongly acquisitive man not only to grab money and power, but to escape much criticism by leaving the district where he made his wealth. He appears in the South as an old English gentleman, pleading for 'our good old English ways'. Going farther than this, we can point out that nowadays the rich man may not even be promoted into the ruling class. He and his female relations may merely become passive members of inner rings of big business and high finance.

Since Mr Priestley's views concern the town-bred money-maker who finds protective colouration among country people, we are left wondering about his opinions of the latter. They are apt to be mute, because talking is usually not their job, still less writing. To

a social psychologist looking for trustworthy information, this is disappointing, for in different parts of England social stratification might present either the same pattern with varying emphasis upon certain parts (the relative widths of the stripes, so to speak, may change) or a different pattern. The economic and political details are obtainable; indeed, many are known. But for the psychological facts we depend at present upon our individual experiences, casual communications and the selective insight of novelists and playwrights. Yet questions about social stratification could be as useful as those which served so well in Middletown, Marienthal and Metrop. Few English people are really unaware of their class, though on some occasions they may act as if they are 'the people' and class is irrelevant.

Miss Margery Allingham's assertion in *The Oaken Heart*, a description of how an Essex village 'took' the first years of World War II, may be true of many an Englishman. When he 'gets on', he hardly ever stays a rich member of the proletariat, but submits himself and his family to the discomforts of a way of life which he thinks is better and grander than his old one. He want to be equal to what he thinks, rightly or wrongly, to be the best. He expects the upper classes to keep the code of behaviour, possessions and manners *up*, so that his ascent may be the higher. He knows well that the journey may take a generation or two, for how many of our best families are very old? We are reminded that some goals exert an enormous pull upon the social climber, though we are not told what they are. Are they always the same? Power, a hope of extinction of inferiority-complexes, of more intellectual pleasures and gentler ways of life for children and grandchildren? And what, in this context, is the meaning of gentle? To this question we will return.

Mrs Wootton evaluates the pulling power of some peaks above the social climber's head. She writes from the standpoint of an economist rather than of the social psychologist; yet *End Social Inequality: A Programme for Ordinary People* illustrates the common fields of these two studies. The opening pages might have been written by a social psychologist (new style). She points out that modern English class differences are, in one sense, very pronounced, while in another they are not. In the first sense, they prevent many people from doing things which it is open to others to do, and cause those who choose to behave as if ours was a classless society to appear highly eccentric or even ridiculous. Yet to avoid the accusation of bad taste one must adapt oneself to the caste system while pretending that it does not exist.

English class-privilege consists of two distinct but related systems. The first she calls opportunity-class; the second social (or snob[2]) class. They must be discussed separately, because, although their foundations have much in common, one might survive even if the other were abolished.

Opportunity-class is simply an unequal distribution of opportunity. There is nothing particularly subtle or evasive about it: at this stage its study is probably a matter for economists and for those medical practitioners sensitive to political problems, as she shows by her first example: the opportunity to go on living in this world, once you have been born. Here we meet our first great inequality. In *Poverty and Population*, Professor Richard Titmuss contrasted the death-rates prevailing in the Home Counties (excluding London) with those in Wales and in a largely industrial area, comprising Durham, Northumberland, Cumberland, Westmorland, Yorkshire, Cheshire and Lancashire. The results showed that if deaths in these latter districts (age for age) could be brought down to the level prevailing in the Home Counties, 53,951 people who actually did die in a single year (1936) would not have done so, This unnecessary loss of life actually exceeds the rate of recorded air-raid casualties during the period of intensive bombing between September 1940 and February 1941. She cites as further examples of opportunity-class influence the choosing, from an arbitrarily restricted group of the population, of persons to hold positions of power and responsibility and to enjoy skilled and interesting jobs.

Snob-class is the system which confers on certain groups of people a *general* (not functional) social prestige It precludes normal social intercourse between those who are high up and those who are low down in this scale of esteem. Snob-class is not the individual distinction, 'glamour' or halo which may surround a film star, boxing champion, explorer, scientist or even philosopher, spilling over, so to speak, from the circle of experience or performance in which they have proved their excellence. Nor is it the prestige which properly accompanies the superior authority exercised by people who occupy positions of power: the respect paid, for instance, to the policeman who controls traffic at a busy crossing.

For this prestige vanishes almost completely when the functionary is off duty – especially when in mufti. In contrast, snob-class privilege gives prestige which is generalised and irrelevant to personal merits and accomplishments. Furthermore it extends to the family of the person enjoying it. Such a family often contains members whose claims to esteem, reckoned on any other scale, would be dim to vanishing-point.

Snob-class, like opportunity-class, is based mainly upon titles, money and education. But its stratification contains some subtleties of its own, e.g. the ownership of landed estates carries a peculiar distinction and the performance of manual work a corresponding stigma.

To sum up, this two-class division in English society means that public life is administered by people who, quite literally, know next to nothing at first hand about the life of the public, are not even conscious of their own ignorance, and tacitly assume that they are typical English men and women. This is not class-consciousness, but class-*un*consciousness.

This statement was written in 1942. I have not altered it, since the present book (*English Social Differences*) and Professor David V. Glass's studies of social mobility in Britain offer data which allow the reader to judge how far matters have changed.

Let us return to consider, as objectively as possible, some of the psychological problems of social stratification.

English social strata have manifest visible and audible signs. To a blind man there are also tactual signs, for example, the soft hands of the leisured. To people with sensitive noses, class may have olfactory divisions. The use of scents – these, too, are socially stratified – suggests that people might be classed as naturally or artificially odorous and unnaturally inodorous.

We will confine ourselves here to the visible and audible signs. The latter are the more important, upper-class speech being much more difficult to acquire than upper-class clothes. In the following description there is no suggestion that one way of speaking is better, more beautiful or more effective than another, though most people have strong prejudices about this subject, as the B.B.C. knows to its sorrow, or amusement.

Social differences are shown by the choice of words and phrases. Each class uses characteristic terms of approval, disapproval, intimacy, endearment, enthusiasm and boredom, and each ignores or ridicules the choice of others. Euphemisms, too, are class labels. Perhaps most of these are used by the middle (especially the lower-middle) class, for reasons obvious in a society permitting vertical movement. Of the phrases which especially indicate social strata, the emollient group are the most striking. A good diplomat has many in stock, and knows when to produce the right one. They are used to begin, interrupt or end a conversation, to refuse or decline a request, suggestion or invitation; to correct a misstatement, to oppose a proposition without rancour, to adduce incontrovertible if unwelcome evidence, to accost a stranger, encourage a shy person, or reprove a backslider. They are learnt after birth, and are not, as a few appear to believe, congenital.

Some readers, interpreting this as an oblique reference to a famous 'North–South' difference, may assert that the Southerner uses many emollient phrases which the Northerner would scorn to employ even if he knew them (which he often does); that they are a sign of insincerity, if not of downright craftiness. Many assertions about North and South 'accent' imply that one is sincerer than the other. Yet 'The North' is a big place. It includes Yorkshire, Lancashire, Northumbria and the Lake District. In all these regions a forthrightness of speech is heard, but dwellers in the North easily identify different varieties and degrees of it. Independent straightforwardness differs, for example, from truculence. And it is fair to record that in the last fifty years when a Northerner became wealthy he often sent his children to schools south of the Trent. These pupils, now grown up, form a powerful section of the middle-aged 'upper-class' in the North. They frequently use emollient phrases, though their accents still may be (they usually are not) pure 'North-country'. This is an example of the difficulty of separating 'geographical' from 'social' dialects, since the public schools spread the latter all over England.

About phrases which mark off social classes I will say little, referring the reader to Professor A. Lloyd James's *The Broadcast Word* and our *Spoken Language*, and to my own *Voice and Personality* and *The Psychology of Effective Speaking*. Most phoneticians are more interested in the geographical and historical differences of spoken sound than in the social stratification of language, though Mr George Bernard Shaw zestfully attacked this theme in *Pygmalion* and the preface to it. He reopened the subject in an introduction

to Professor Richard Albert Wilson's *The Miraculous Birth of Language*.

One social group may inaccurately attribute 'bad' speech to another. American journalists, when they 'quote' English working-class speech, are apt to write it without aspirates, yet in many parts of industrial England the 'h' is seldom dropped. With such warnings in mind, it may be said that difference in vowel-pronunciation is an important mark of social class. The Cockney modifies all the English vowels, usually transforming them into diphthongs. Cockney radio-impersonations represent various types – even different generations – of this 'district-dialect'. It is interesting that modern upper-class fashions in pronunciation frequently approach, though without quite touching, the Cockney language; e.g. 'Go home', pronounced by a modern actor, may be almost indistinguishable from a coster's speech. Both, Mr Shaw remarks, may call the sun the san and a rose a rah-ooz. A B.B.C. announcer has been heard to say 'Tao-Keaow' But the Cockney pronunciation of 'Bill', 'milk', 'field' and 'towel' has not been adopted by fashionable English people.

Of speaking, considered as a means of social communication, 'speech-melody' is perhaps the subtlest characteristic. By varying it emotional relationships are established with others who understand our language. This proviso must be remembered. Since intimacy or 'distance', friendliness or enmity, interest or boredom, command or prayer, deference or disrespect are expressed by speech-melody, it is not surprising that social differences are emphasised by the way in which words go 'up and down'. This is well known to self-constituted judges of good manners, especially those who, on the basis of a brief interview, recommend candidates for posts.

The auditory 'aspect' of speech is perhaps more important than the visual accompaniments, facial expression, gesture and posture. Possibly the comparative absence of facial expression and gesture from the speech-behaviour of the English 'ruling classes' is a sign of social stratification observed particularly in officers in the older fighting services and in some school prefects. Within the different geographical dialects of England there are social layers. Lancashire provides many different ways of speaking. Some of those reflect social gradations. The Londoners in the first act of *Pygmalion* all speak differently. Until it is widely recognised that many variations are social rather than geographical, some of the chief causes of English social differentiation will be overlooked.

Concerning clothes, in peace-time, as Dr J. C. Flugel has shown, there are two continuous opposed waves of influence, towards and away from standardisation. Some expensive clothes are copied and then mass-produced; when this happens, further designs, resembling these in shape or shade, yet clearly distinguishable from the popular model, are produced in limited quantities by 'exclusive' firms. Perhaps some inexpensive clothes for men are shaped to look cheap, for the same firm may sell both these and dearer suits. This seems a deliberate attempt, encouraged by public support, to perpetuate class-difference. Similarly, speakers who in standard English exhort others to adhere to their dialects may not be keen egalitarians.

The psychology of eating and drinking (what is eaten, how, when, where, and with whom) involves consideration of social differences. Some of these immediately become clear to those who faced the problems of communal feeding. Certain hotels, restaurants and public-houses are recognised as the resorts of different classes. This fact is reflected in their prices, but not simply, at least in London. A large hotel may contain restaurants, and a public-house bars, socially graded.

Table manners are admitted class-signs, a fact which evacuation in war-time underlined. Foreign observers have noted the sociological importance during the last century of our tea-shops and Corner Houses, offering to thousands opportunities to eat in public which formerly did not exist for them – since many would never willingly enter a public-house and could not afford the nineteenth-century hotel or restaurant.

Differences in social status arising from the possession of property have been often described by novelists: John Galsworthy succeeded so well in presenting the Englishman as a Man of Property that many Continental readers firmly believed the Forsytes to be a common English type. To some extent they were; for centuries English law and politics have been very kind to the landowner, and offences against property are still severely punished. This fact complicates the treatment of juvenile delinquency, especially in war-time.

The type of house in which a person lives, its locality – in London, and presumably elsewhere, long streets may have a 'right'

and 'wrong' end – its furnishings and pictures, and its garden, if any, give him social standing. So does the possession of an automobile (the make and date often being taken into account), a telephone, a radio and a television set. At least one social survey has been wrecked because the people canvassed were telephone-renters, whose names were thus easily obtainable, but who, because of this fact, were all above a certain income-line.

No Englishman needs reminding that in the matter of education there are two Englands. The social and economic differences between our types of educational 'facilities' appear to be the wonder of the world. At present they are being discussed so widely that a special chapter would be required even to summarise the conflicting arguments. Without taking sides in this controversy, it may be remarked that though the public schools have virtues which an ideal system of education could well retain, to a democrat their most undesirable feature is the deliberate segregation of their pupils, during their formative years, from the rest of the community. It is possible for an English boy at the age of eight to go to a residential preparatory school, then to a public school, afterwards to a university, and on graduation to join the staff of another (or the same) public school; to live to a ripe old age, having taught thousands of boys, yet never having been in the outside world except for holidays. Even those will probably have been spent among people of his own class. Of all the factors which, during the last half-century, have perpetuated and increased social stratification, the public schools are in the front rank.

The type of religion which the Englishman professes often reflects his social standing. Stratification is very obvious in the Church of England. Distinctions between 'high' and 'low' church, though usually assumed to refer to differences of preference for ritual, are often social. In many parts of England there is still a social difference between 'church' and 'chapel'. The social status of English Roman Catholics in general is undefined, but among themselves they are not unstratified.

In this space, only the general directions of arguments put forward today to justify or to condemn social stratification can be indicated. The case for it is supported by asserting that if, for centuries, there is freedom for selective mating (obviously the poor, lacking chances to travel, and to meet many different people, have less freedom), the offspring are likely on the whole to be healthier, stronger, more beautiful and more intelligent. A leisured class will encourage and defend culture, appreciating its advantages. It is urged that if one 'knows one's station', with its privileges and corresponding duties, this promotes easy working of the social scheme. A hereditary aristocracy, it is argued, tends towards tolerance (arising from knowledge that individual differences are great), 'intuitive' leadership, and a sentiment of responsibility for others, developed by generations of feudalism; *noblesse oblige*. The force of these arguments, however, assuming their validity, depends upon the privileged stratum retaining its characteristics. During the last century the social layers from which our ruling class is drawn have not remained constant.

Many have criticised the above claims, or have accepted some of them, commenting that such assertions could be made of any privileged class, and that, to use an Irishism, everyone ought to be privileged, if this means to have opportunities for leisure, travel, training in social responsibility, and where desirable leadership. In *The Dangers of Being a Gentleman*, Professor Harold J. Laski described the ideal of gentlemanliness. It may be summarised here.

A gentleman is, rather than does. He is interested in nothing in a professional way. He is allowed to cultivate hobbies, even eccentricities, but must not practise a vocation. He must know how to ride and shoot and cast a fly. He should have relatives in the army and navy and at least one connection in the diplomatic service. But there are weaknesses in the English gentleman's ability to rule us today. He usually knows nothing of political economy and less about how foreign countries are governed. He does not respect learning and prefers 'sport'. The problem set for society is not the virtues of the type so much as its adequacy for its function, and here grave difficulties arise. He refuses to consider sufficiently the wants of the customer, who must buy, not the thing he desires but the thing the English gentleman wants to sell. He attends inadequately to technological development. Disbelieving in the necessity of large-scale production in the modern world, he is passionately devoted to excessive secrecy, both in finance and method of production. He has an incurable and widespread nepotism in appointment, discounting ability and relying upon a mystic entity called 'character', which means, in a gentleman's mouth, the qualities he traditionally possesses himself. His lack of imagination and the

narrowness of his social loyalties have ranged against him one of the fundamental estates of the realm. He is incapable of that imaginative realism which admits that this is a new world to which he must adjust himself and his institutions, that every privilege he formerly took as of right he can now retain only by offering proof that it is directly relevant to social welfare.

In short, 'breeding' has often worked well, but for types of social pattern which no longer exist. Sometimes, however, it worked badly even for them.

But all classes, not one alone, ought to have leisure, in which culture can be attained. Tolerance – and intolerance – are not the prerogatives of any one class. Both these characteristics can be acquired, especially when taught deliberately by example and precept. During the last fifty years the upper classes have often failed to shoulder responsibility. Finally, culture is acquired, not innate. The baby, even if born in the purple, is lamentably uncultured.

The case against privilege has been argued by Professor Lancelot Hogben in *Dangerous Thoughts* and by Mrs Barbara Wootton in *End Social Inequality*.

The views about English social stratification held by readers of newspapers and novels (it is from these sources, as well as from parental teaching and example and the social environment in school, work, pastimes and sports that the average person's opinions come) may truly represent certain facts but not others. It needs no deep insight to see that in England yachting is an upper-class pastime while pigeon-flying is not; that Rugby football is socially in 'higher' esteem than the Association code. Other games may rise and fall rapidly. Until recent times, darts was a lower-class game. It was embraced by the upper classes when the Queen Mother, visiting a welfare centre at Slough, happened to take up a dart and throw it at a target. This event, immediately followed by the 'pushing' of darts by the B.B.C. who, unopposed by brewers, sedulously linked up the idea of the game with that of public-houses, made darts popular with all classes for a time.

Some aspects of our social structure, however, may have changed imperceptibly in the last few decades. Several of these were studied in 1939 and 1940 by Mrs Wootton at an advanced tutorial class, of which she was the teacher. Their object was to form

> as realistic a picture as possible of the class structure of our society, and to test against this the validity of opinions commonly held about the meaning and relationship of classes. In particular to discover what, if any, are the realities that correspond today to the terms 'Working Class' and 'Ruling Class'.

They distinguish two senses in which the term 'class' is commonly (if unthinkingly) used:

> (1) A group of people who have a 'common economic interest', i.e. whose common economic interests outweigh their economic conflicts.
>
> (2) A social group, i.e. a group of people who mingle in ordinary social intercourse and are set in a hierarchy whose order is generally recognised, so that they are debarred by convention from so mingling with groups above or below them in the social scale.

They divided 'Economic Class' into

> (1) Groups with a subjective economic interest only, who imagine they are economically united, but in fact are not.
>
> (2) Groups with an objective economic interest only, a real bond of which they are unconscious.
>
> (3) Groups with an objective and subjective interest.

'Social Class' they defined as

> groups of people who mingle with each other in normal social intercourse (e.g. entertain each other in their homes) without restraint but who are conventionally debarred from so mingling with other groups who are socially inferior or superior according to a generally recognised hierarchy. This hierarchy is a scale of prestige; prestige which if related to the individual's function in society (occupation) extends beyond the function both during non-exercise of the function and into retirement. (For example, a factory manager is accustomed to be called 'Sir' by his workmen not only in the shop but at the sports club and the bar parlour.) The prestige of an individual extends normally to his family and dependants.

After considerable examination and discussion, they decided that the quantitative prestige-differences which are the basis of social

stratification in Great Britain today depend on wealth, birth, education and occupation. Weighing these four sources of prestige against each other, they formed the opinion that the greatest of these is wealth, and adduce evidence for this. They continue:

> If the foregoing were only what we believed to be the nature of the social hierarchy in Great Britain it could reasonably be dismissed as pure conjecture. It coincides, however, with what we identified as equally important in our study of particular individuals.

They examined certain characteristics of the people who rule Britain, i.e. those who possess political and administrative power. Concerning the alleged 'ruling class' their conclusions were:

> (1) Family, Education, Occupation and Recreations of men in ruling positions are by no means typical of an average group of the population.
>
> (2) The following attributes appear in much higher frequency amongst these rulers than amongst other people:
>
> *Family* – Titled and Professional.
> *Education* – Public and Grammar School, University education.
> *Occupation* – None or Professional.
> *Clubs* – Carlton and Brooks.
> *Recreations* – Shooting and Golf.
>
> These attributes can be called ruling-class attributes, because people who possess them have a more than average chance of acquiring ruling power.
>
> (3) The ruling class is primarily a social group, maintaining its exclusiveness and unity primarily by the possession of wealth, with the other characteristics studied playing only a secondary part.
>
> (4) The possession of wealth and the wish to keep it is the one objective economic interest which outweighs the many economic conflicts found in the ruling class. There is, therefore, a clear objective basis for political rather than for industrial unity.
>
> (5) The ruling class is not so much a class of employers set against all employees, as it is a class of rich people set against poor.
>
> (6) A large percentage of those people who attain ruling power without coming from the ruling class are Trade Union leaders.
>
> (7) People who are not ruling class are seldom found in any but *elected* positions of power.
>
> (8) As long as attendance at expensive schools and universities is the training needed for administrative power, there will still be a ruling class, for poor people will be excluded from the necessary training.
>
> (9) As long as the people who rule are paid a salary which is very much higher than the average salary, their social habits will cut them off from the average man.

Concerning the working classes they concluded:

> (1) If there is a working class, there is no common agreement on its definition.
>
> (2) None of the definitions offered pointed to a common objective economic interest.
>
> (3) Many workers might feel a common interest in that they all worked, or were important to society, or that their work was misapplied.
>
> (4) The significant class struggle is not between workers and employers, but between those who work and those who do not.
>
> (5) Within the widest definition of working class there are distinct social groups. These correspond to some extent with a subjective economic interest. Social stratification overrides economic unity.
>
> (6) The general belief in the existence of a working class is due to the exclusion of certain groups from the ruling class.
>
> (7) The factors making for the exclusion are poverty, lowly birth, humble occupation and lack of education.
>
> (8) The most important factor in class exclusion and class unity is poverty, not the fact of being a worker.
>
> (9) The belief in the existence of a working class is justified, but that class would better be called the poor.

During recent years many events have given us chances to note the effects of social stratification. Evacuation of children and mothers from dangerous to safe areas offered one opportunity, and many publications about the impact of class upon class appeared. Everywhere one fact emerges: difficulties in adjustment to strange

social situations caused by the manners and customs of their hosts were very much less for the younger children. Some will regard this as a blinding glimpse of the obvious; yet it should be recorded and emphasised because if we wish to decrease class-consciousness it shows one easy way to begin. Day nurseries, nursery schools and nursery classes offer a chance of diminishing social 'distance'. It would also be helpful to discover the ages and conditions which make awareness of class-difference most intense.

One might add that the difficulties of evacuee children in meeting novel social demands were not always due to class-difference. In one report, kindly sent me by the headmaster of a city school whose pupils went to a seaside town (names are omitted for obvious reasons), the boys were all 'lower middle-class'. There were few social distinctions between them. The hosts appeared to be of a class similar to that of the pupils, yet real difficulties arose, due to local usages and preferences, or to age disparity between hosts and guests.

The evacuation reports seldom analyse the subjects of disagreement and friction, though many are named and described. They may be divided into avoidable and unavoidable – e.g. a person who has regarded solitude in a bedroom, or a quiet place in which to read and write, as necessities, may, if he entertains evacuees, never get used to the continual company of unchosen guests. Dislike for close contact with people less clean than oneself may seem just as deep-rooted, yet many lose or dissemble it in army or camp, and during the 1940–1 winter, shelter-life rapidly decreased the cleanliness of thousands, not all of whom regarded this as an unmitigated evil. There are class prejudices due to unreasoned dislikes, traceable to early conditioning, or to personal complexes. People who have overcome this weakness may still wince at a few varieties of transatlantic speech, though tolerating some of its ingredients (e.g. the hard 'r') in West-Country dialect. To some ears, Aberdonian 'pure' English sounds mincing and finicky. Some are affronted by all 'upper-class' accents; this attitude may arise from ambivalent inferiority-complexes. Many travelled English, readily ignoring 'foreigners'' unusual methods of handling knife and fork, are disdainful of them in fellow-countrymen of a lower social class. Others, gratified when an American says he is pleased to meet them, are amused at an Englishman's use of this phrase. The employment of circumlocutions and of long or 'elegant' words indicates class, and many English are contemptuous towards compatriots who employ unnecessary or 'wrong' ones, but not towards English-speaking foreigners who do so.

Social psychologists have seldom or never studied differences between English codes of manners. Books of etiquette exist, codifying some manners of the upper classes, implying that they set the standard for everyone. Yet these books 'date', and after the war the best of them may have to be rewritten. Questions to which a standard guide to manners might give out-of-date answers concern the use of alcoholic drinks, tobacco and cosmetics (especially regarding time and place), and the modes of address and behaviour of one sex to the other. In the early years of World War II, letters to the newspapers suggested that an A.T.S. private would resent a reproof less if it came from a man officer than from a woman, an assertion which starts a complicated train of thought.

Some may believe that there are fundamental, unchangeable principles of civilised behaviour. Though this may be question-begging, since it is difficult to define 'civilised' in a way which would suit everyone, we may assume, in our own culture-pattern at least, certain conventions which most people consider essential. Interest might therefore be taken in some articles concerning the proposed democratisation of the Diplomatic and Civil Services, which appeared in the *Manchester Guardian*. In the issue of 30 June 1941, Mr Thomas Jones discussed some proposed changes in the Foreign Office, and the possibility of recruiting the Diplomatic Corps in ways which would not prejudice a candidate without private means. In his view, there was little hope of getting desirable men from the classes usually termed lower.

In a letter to this newspaper (17 July) Professor Harold J. Laski examined some of Mr Jones's arguments. One was the value of ' "contacts" due to the associations of well-travelled families'. Professor Laski replied, 'The underlying assumption of this argument is the maintenance of an unequal society in Europe in which only the well-endowed will have the opportunity of foreign travel.' Mr Jones had written, 'The grace and ease of cultivated conversation are a lubricating accomplishment normally more often found in the youth of Eton than of Plumstead.' But, answered Professor Laski, 'Most of the youth of Plumstead do not have the advantages of the youth of Eton. Those of them who, through scholarships,

have the advantage of an equivalent education are not, in my 25 years' experience as a university teacher, less capable of the "grace and ease" of which he speaks. And in general they are more open to new impressions and to the significance of new experience.' To Mr Jones's assertion 'There is no presumption in favour of the products of our elementary schools being inherently better qualified for diplomacy than the youth of the great public schools', the reply was made 'That is not the question. What disturbs the critics is that the system now in operation should assume that over 60 per cent of the natural diplomatic ability of this country should, in the last 75 years, have been concentrated in some half-dozen schools. It has not been thus in the case of other faculties. Is it not possible that diplomatic ability is only so limited because the system has not responded to changes in the structure of European society?'

Passing to the Civil Services, Professor Laski wrote, 'The main traditions of the service are hostile to initiative, risk-taking, large-scale experiment. Those, for the most part, who climb up are the men who impress their superiors as "sound and safe" and likely to conform to the tradition; this is especially true of those who are selected for promotion from the lower grades. In 15 years' experience of the Civil Service Tribunal I have met many men of first-class ability in the executive and clerical classes who have stagnated there. They have not been lifted out of their grades because they do not accept the unstated assumptions which shape the central tradition of the administrative class.'

A few years ago few would have regarded such a difference of expert opinion as a matter for psychologists. Social gradations exist; they have made England what she is; why drag them into the limelight? Nowadays, however, it is reasonable to inquire whether the characteristics discussed above can be analysed and their development explained.

In the *Manchester Guardian* of 5 August 1941, the present writer suggested that the psychological constituents of diplomatic and democratic manners might be examined. Especially important are intonation, speech-melody and choice of words and phrases suited to particular occasions, such as committee meetings or debates. The Englishman's reserve is less mysterious to, and perhaps less respected by, 'foreigners' than it used to be. Ability to converse and discuss do not result from 'instinct' or selective breeding. If certain classes discuss better than others, the reasons ought to be discovered.

Few psychologists, for example, have tried to interpret data obtained by phoneticians. Many of their problems, e.g. prejudices concerning standard English and dialects, are matters for social psychology. To treat them thus, however, offends the taboo already mentioned.

In attempts to support the view that the upper classes are specially fitted for the diplomatic services, it is pointed out that the diplomat's aim is not mere politeness; that knowing 'who is who' and 'what is what' is more important. This assumes that the social knowledge upon which discrimination may be exercised is the prerogative of one class. Critics assert that occasionally, e.g. in 1933 and 1939, our representatives do not seem to have known who is who, and urge that a diplomat should be intelligent, honest, truthful, friendly and know the relevant facts.

To examine and criticise such assertions is necessary if democratisation of the Diplomatic and Civil Services is to be tackled effectively. It presents clearly many thorny points of English social stratification. Two people, one whose family had been connected with the diplomatic services for generations, the other a fervent democrat, would have to sink many prejudices before they could even begin to discuss the question profitably. To an Englishman this may seem natural, requiring no comment, yet in younger countries comment upon our social layers is free and pungent.

The solution of many problems in social psychology needs more knowledge of the sentiments and ideals which characterise different social strata. Often an author's class can be accurately deduced by examining his postulates. On reading attempts made between 1918 and 1939 to explain war in terms of human experience and behaviour, one seems driven to conclude that most of them were conceived from an upper-class standpoint. Underlying them seems to be the assumption that when war is declared, the author will take – or be holding – a commission, and will lead 'his' men. Few, especially those who 'explain' the stirring-up of war in peace-time as due to aggressiveness, ask themselves how much the poorer people know about the nations for whom, after war has broken out, they are exhorted to feel emotions specified by the propagandists.

To conclude, my view, I hope, is clear. In the past, privilege may have exercised good influences in many directions, but during our lifetime its shortcomings have been increasingly obvious. Today it seems to be falling into disfavour all over the world. Yet the difference ought to be clearly recognised between lessening privilege and decreasing tolerance of individual differences. I write this, though believing that recent attempts to retain privilege at all costs were powerful factors in bringing about the last war. If democratisation sets in suddenly and violently, there is a danger that not only the polish of social life but the material upon which during centuries that polish has been achieved, may be destroyed. I would urge the social psychologist to study the more and less desirable social phenomena, both in the upper and the lower classes, to find out how, when the cause was not economic, they have arisen. For gentle manners are not the result of an accident, and it would be a tragedy if we lost them.

A social psychologist who makes this study must expect to be shot at from both sides. Weapons used against him may be not only opposition, but 'good-humoured' ridicule, an effective instrument of attack today. Yet standards of humour have been known to change, even in England.

Notes

1. Sir Cyril Burt comments: '19th-century psychologists, particularly those who described themselves as Philosophical Radicals, were fully aware of the influence of class differences. At the beginning of this century the subject attracted considerable attention from writers like Graham Wallas, L. T. Hobhouse and William McDougall. It came still further to the fore in the early controversies between the eugenists and the environmentalists, and educational psychologists, particularly those influenced by Sully, discussed its relation to the problems of the schools.'

2. I have tried and failed to find a synonym, or even a near-synonym, which describes those who sincerely believe that the English class-system is a good thing.

THE BASIC MODEL

Johann Heinrich von Thünen lived in Germany from 1783 until 1850 and at the age of twenty-seven acquired Tellow, an estate near Rostock in Mecklenburg, which he continued to manage for a further forty years. During the course of his work he amassed a considerable amount of data concerning farm costings and at the same time studied both economics through the works of Adam Smith and scientific farming with Albrecht Thaer. Eventually, in 1826, the results of his experience were incorporated in a publication entitled 'The Isolated State' ('Der isolierte Staat'). Its aim was to discover the laws which govern the prices of agricultural products and the laws by which price variations are translated into patterns of land use. This work was essentially that of a practical farmer, for it was of interest to him to appreciate the financially most rewarding system for conducting his own enterprise. As it is, he in fact produced a method of analysis for rural land use which can be considered at the same time as the first theory of the location of agricultural production.

At the basis of von Thünen's work was the concept of the isolated state, as the title of his publication suggests. He described the nature of this state at the beginning of his opening chapter:[1]

Imagine a very large town, at the centre of a fertile plain which is crossed by no navigable river or canal. Throughout the plain the soil is capable of cultivation and of the same fertility. Far from the town, the plain turns into an uncultivated wilderness which cuts off all communication between this State and the outside world.

There are no other towns on the plain. The central town must therefore supply the rural areas with all manufactured products, and in return it will obtain all its provisions from the surrounding countryside.

Note

1. *Von Thünen's Isolated State*, trans. C. M. Wartenburg, ed. Peter Hall (Pergamon, 1966) p. 7.

The assumption was also made that there is only one mode of transport between town and country and that farmers who supply the central city have no problem in adjusting the use of their resources to any economic condition that might evolve – in fact they act with complete rationality.[1] *It is therefore quite apparent at this stage that several factors which normally contribute to considerable regional diversity in rural land use and agricultural production were assumed either to be constant or not to exist at all.*

From these basic assumptions von Thünen developed his arguments in detail. These were summarised in Andreas Grötewald's paper 'Von Thünen in Retrospect' (1959).[2]

LAND RENT

Within the setting of the *isolierte Staat* von Thünen considered the relationship among (1) distance of farms from the market, (2) prices received by farmers for their products and (3) land rent.

The relationship between the first two factors was very simple. The price which a farmer obtained for a given unit of commodity was equal to its price at the market minus the cost of shipping it to the market. The cost of transportation increased with distance from the market. Thus, any given product – a bushel of rye, for example – represented a greater value to a farmer living within two miles of the market than to a farmer living at a distance of twenty miles.

The third factor, land rent, was defined as the return from the investment in the land.[3] In the *isolierte Staat* land rent depended upon location relative to the market. This can be explained most easily by a history of settlement. When the *isolierte Staat* was small, only a little land around the city was needed to supply it with agricultural products. As settlement expanded and the demand for agricultural products increased, land at a greater distance had to be taken under cultivation. As a result of the increased cost of shipping from this outer zone to the city, market prices for agricultural products had to increase. Farmers in the earlier settled area benefited from such higher prices. Compared to farmers in the more recently settled outer zone, they obtained the differential advantage of lower transportation costs to the market; thus their land provided a rent and increased in desirability and value. Eventually, settlement expanded still further and a third zone,

even more remote from the city, was taken under cultivation. At this stage prices for agricultural products in the city had to increase even further. Consequently, the second zone of settlement began to provide a land rent, and in the first settled area the rent increased.

The formula given below indicates how land rent for any one product can be derived:

$$R = \gamma p - E - \gamma f k.$$

In this formula, R = land rent per acre; E = production expenses per acre, including labour, supplies and equipment; γ = yield in units of commodity per acre; p = market price per unit of commodity; f = freight rate, i.e. the cost of shipping a unit of commodity over the distance of one mile; and k = number of miles from the market.

The value of land was determined by the rent it yielded. Land nearest the city was most valuable because it yielded the highest rent; land near the outer limit of cultivation was of little value because it yielded a low rent.

AGRICULTURAL PRODUCTION

Within the *isolierte Staat* areal differences in land use and agricultural production were the result of three forces. These were (1) the types and quantities of agricultural products needed in the city, (2) the technology employed in the production and transportation of such commodities, and (3) the endeavour of each farmer to maximise his land rent by producing commodities for which the location of his land relative to the market offered the greatest advantage. These three forces created the pattern of land uses shown in Fig. 19.1(*a*).[4] The following is an interpretation of this pattern.

It has already been mentioned that von Thünen relied upon empirical information in defining the commodities needed in the central city and in calculating for such commodities the production expenses, yields and market prices. His model of the *isolierte Staat* reflects the needs of Rostock city dwellers and the state of technology in Mecklenburg well over a hundred years ago. Writers who have since described von Thünen's theory or elaborated upon it have frequently overlooked these circumstances, and such neglect has resulted in false interpretations and unjust criticism of his theory. Naturally, von Thünen took the conditions of his time for granted. Today, however, such conditions must be stated as definite propositions of his theory. One of these is that in the *isolierte Staat* there was only one form of transportation, the horse and wagon, operated by the farmer at his own expense. There were no multiple freight rates depending upon the commodities shipped. Refrigeration in transit had not yet been invented.

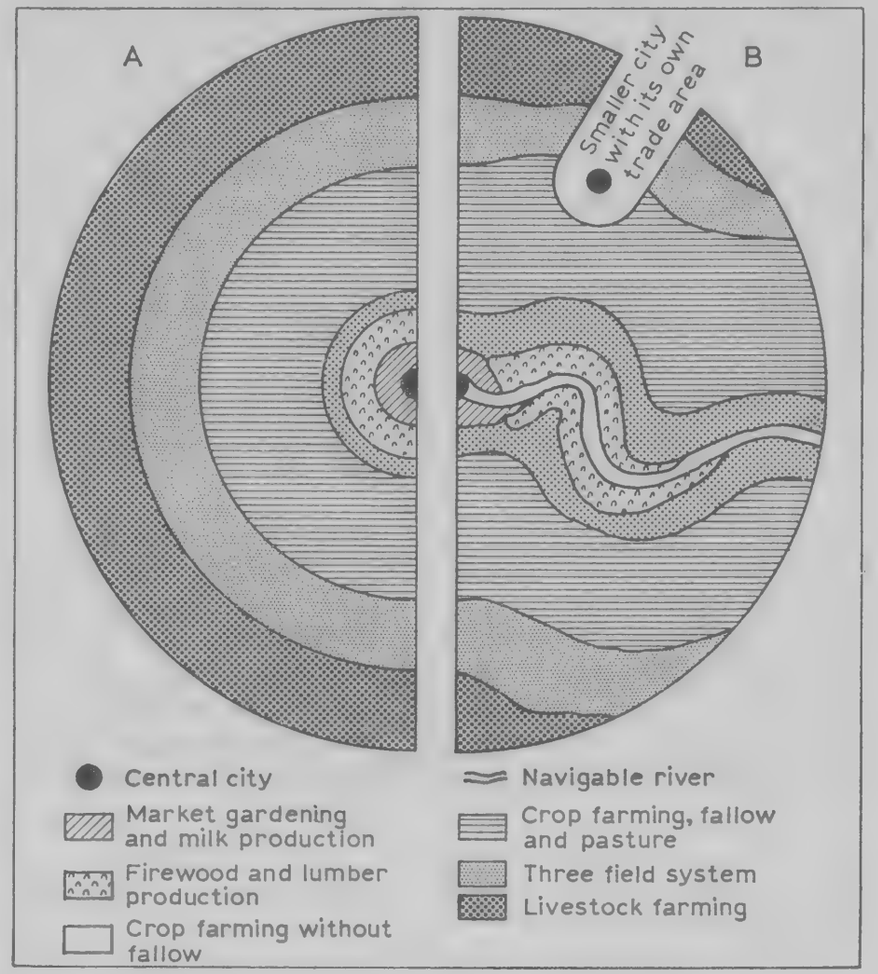


Fig. 19.1 The location of agricultural production after von Thünen: (*a*) The *isolierte Staat*; (*b*) modified conditions

Under such conditions the land nearest the city was used for the production of fresh milk and vegetables. Here cattle were kept in stables throughout the year. The fertility of the land was maintained through manuring, and if necessary additional manure was bought from teamsters in the city. The reason for the existence of this zone so close to the city is to be found in the state of technology. The slow speed of transportation and the absence of refrigeration made it impossible, or at least precarious, to ship fresh milk and vegetables over greater distances. In the location of this particular zone the cost of transportation was not a pertinent consideration. Since fresh milk and vegetables were required in the city, prices for them had to be so high that no other use of the land would have yielded a higher rent.[5]

The second belt was used for the production of wood. Von Thünen has shown, on the basis of his empirical data, that forestry yielded a higher rent on this land than any other use. The following is a simplified version of his argument.[6]

Forest land produced not only lumber but, more important, it was the source of firewood required in the city. The physical volume of all wood obtained from an acre was quite large; let us assume that it was three times as large as that from an acre of product D. In addition, let us assume that for each commodity the production expenses and yields were the same in all parts of the *isolierte Staat*.

Fig. 19.2 illustrates the competition for the land between wood (thick lines) and product D (thin lines). The dashed lines (Yp) represent the prices of produce per acre at the market. The dotted lines (E) show the production expenses per acre. The solid lines ($Yp - Yfk$) represent the market prices of produce per acre minus the cost of shipping them to the market, i.e. the per acre values of the respective commodities at the site where they were produced. Obviously, these values decreased with distance from the market. In our example the rate of decrease per unit of distance was three times as great for wood as for product D. All landowners in the *isolierte Staat* selected the land use which would give them the highest land rent (R). Fig. 19.2 shows that wood yielded a greater rent up to the distance of k_2; beyond that the rent was higher for product D. Thus, wood was produced closer to the market than product D.

Some authors have stated as a general law that, according to

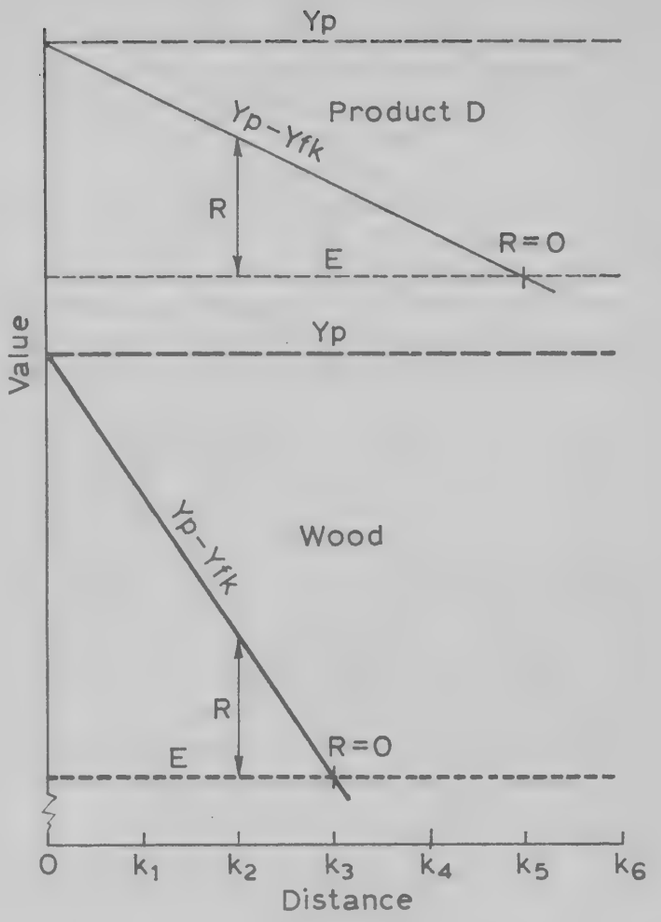


Fig. 19.2

von Thünen, the product with the higher yield per acre must be produced closer to the market. Von Thünen had indeed arrived at this conclusion and stated it in his book.[7] But it is misleading to repeat this statement outside its proper context. It held true only if the competing commodities were equally resistant to spoilage and if the higher cost of shipping the more bulky product could not be compensated by lower freight rates or by a reduction in the production expenses.

According to von Thünen the production expenses for wood were fairly equal in all parts of the *isolierte Staat*. This, however,

was not so for all commodities. His model of the *isolierte Staat* (Fig. 19.1(*a*)) shows behind the forest belt three zones where rye was an important market product. The essential difference among these three zones was in the intensity of cultivation. With increasing distance from the market the production expenses for rye were lowered with a consequent reduction in yields.

In the first of the three zones farmers followed a six-year crop rotation. Rye occupied one-third of the land; the other two-thirds were planted to potatoes, barley, clover and vetch.[8] Potatoes, like rye, were sold in the city. Vetch was used as a soil builder and for green fodder during the summer. Cattle were kept in stables most of their lives. The soil fertility was maintained through manuring. There was no fallow land.

In the next zone the land was used somewhat less intensively. Farmers applied a seven-year crop rotation wherein rye occupied only one-seventh of the land; it was followed by one year of barley, one year of oats, three years of pasture and one year of fallow.[9] The market products consisted of rye, butter, cheese and possibly live animals to be slaughtered in the city. Such products did not spoil as quickly as fresh milk and vegetables and could be produced at a considerably greater distance from the market.

In the most distant of the three rye-supplying zones farmers followed the three-field system. This was a rotation scheme wherein roughly one-third of the land was used for field crops, another third for pasture and the remainder left fallow.

Beyond this zone, in the area of livestock farming, the distance to the market was so great that even the most extensive and cost-saving method of rye production could not have yielded a higher rent than the production of butter, cheese or live animals. Consequently, the production of rye was limited to the demand on the farm. Only animal products were marketed.

Fig. 19.3 illustrates the effect of decreasing production expenses and yields upon the areal extent of rye cultivation. Rye A (thick lines) represents rye cultivation as it was described by von Thünen; production expenses and yields decreased with increasing distance from the market. Rye B (thin lines) is a hypothetical projection of rye cultivation as it was carried on near the market into areas at some distance from the market. At the market the rent for rye A and rye B was the same. From there outward, however, the rent decreased less steeply for rye A than for rye B. At the distance of k_4 the rent for rye B would have been zero whereas for rye A it was still positive. Thus, through less intensive cultivation the rent which rye yielded at greater distances from the market was increased, and rye could compete with other products over a wider area.

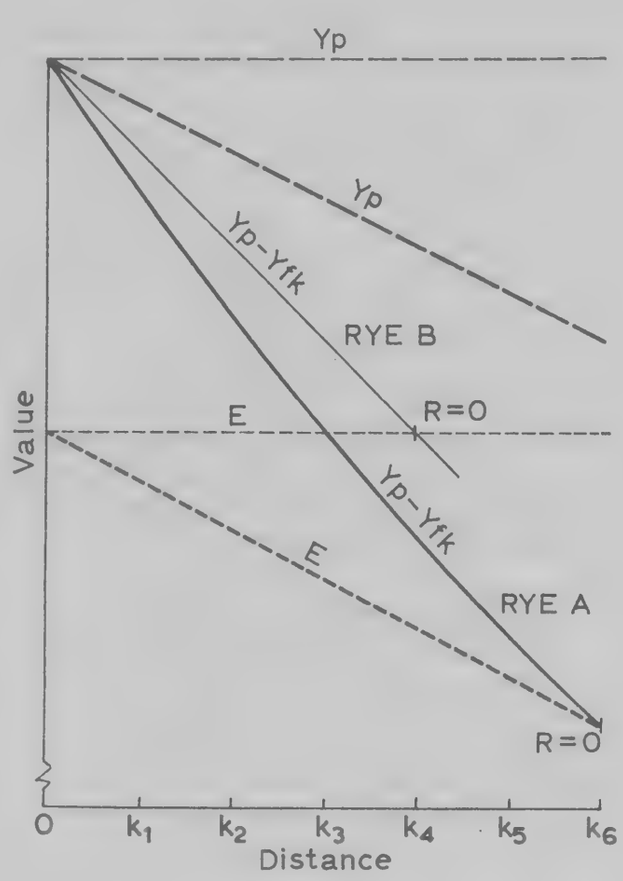


Fig. 19.3

On the basis of this argument some writers have stated as a general law that, according to von Thünen, the intensity of cultivation must decrease with increasing distance from the market. Such statements are very irritating if quoted outside their proper context. As is shown in Fig. 19.2, the production expenses for wood

were far lower than for product D. Yet wood, the less intensively cultivated product, was grown closer to the market than product D. Even in the case of a single product, such as rye, this law is subject to an important qualification. A reduction in the production expenses is feasible only if the consequent reduction in yields is not so great as to offset the savings accomplished by the reduction in production expenses.

Notes

1. Rationality supposes that the entrepreneur, the farmer, pursues the goal of maximum profit and is possessed of all the necessary knowledge to that end.
2. *Economic Geography*, vol. 35, pp. 348–53.
3. The discussion of land rent here is not paraphrased directly after von Thünen's book. But the definitions given are synonymous with von Thünen's. See Thünen, *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, 3rd ed. (Berlin, 1875) pt 1, pp. 14–16, 227–9, and 350–3.
4. Fig. 19.1 was copied from ibid., pt 1, pp. 390 and 391. Some of the terms used by von Thünen to describe types of farming have no equivalent in the English language. An attempt had been made to translate their meaning as closely as possible.
5. Ibid., pt 1, p. 3.
6. Ibid., pt 1, pp. 172–84.
7. Ibid., pt 1, p. 2.
8. Ibid., pt 1, p. 90.
9. Ibid., pt 1, p. 89.

SOME REAL-WORLD VARIABLES

It has to be recognised that the conditions found in the isolated state are nowhere found in practice. Von Thünen recognised this and in Part II of his book he showed how the concentric zoning around the central market could be altered as shown on the right-hand side of Grötewald's Fig. 19.1.

To begin with, he abandoned the notion of a single city by discussing the modified zones of production which would appear if a subsidiary city were introduced. He did not, however, press the matter to the point at which real-world complexity would be introduced by postulating several towns of more or less equal importance with production zones intermingling and modifying each other.[1] *In any case it might be argued that this notion can be considered essentially a modern one and that in von Thünen's time central markets were more widely separated in relation to the then existing transport technology.*

It will be recalled that in his initial formulation von Thünen assumed that not only was the cost of transport proportional to distance but that there was one transport mode only. He later relaxed the second of these assumptions by showing what happens when an improved means of transport, such as a navigable river, is introduced. Through the use of the river, costs were less than those incurred through the use of his original mode, overland cart. Although once again the cost of transport along such a route was assumed to be proportional to distance travelled, a journey which used two or more transport modes could involve different costs for the same distance. The final cost would depend on the distance travelled by each mode. The river, as a means of better communication, was assumed ro run through the city as in Grotewald's Fig. 19.1. The net effect of this, as can be seen, was to expand the zones of production along the line of the river while contracting them elsewhere.

In a real-world situation, however, whatever modes of transport are used (and it must be remembered that in von Thünen's time cart or waterway would have been the main alternatives), the cost of transport is not, as he assumed, proportional to distance. To illustrate this point, the work done by J. W. Alexander on the nature of railway freight rates for agricultural and other commodities may be quoted. In 1963 he summarised two of his papers on this subject[2] *in his book 'Economic Geography'*[3] *as follows:*

One might suppose that the price paid by customers for shipping freight would be commensurate with mileage. Although no world-wide study has been made to test the validity of this assumption, studies within the United States reveal that freight rates themselves are remarkably variable and that they do not vary consistently with either length or direction of shipment. Five aspects of freight rates illustrate this point: freight-rate territories, the rate-group principle, class rates, commodity rates and in-transit rates.

1. Freight-rate territories

The freight rates charged by railways in the United States are regulated by the Interstate Commerce Commission. This agency divides the nation into five regions (see Fig. 19.4). For several decades before 1937, when this map was constructed, the Commission authorised lower rates in the Eastern (or Official) Territory than anywhere else. As the map shows, prices were highest in the Mountain-Pacific Region, averaging 90 per cent more than those in Eastern Territory. The Interstate Commerce Commission defended this policy on the grounds that traffic was so heavy in

Fig. 19.4 United States freight-rate territories

Source: U.S. Government, 75th Congress, First Session, *The Inter-Territorial Freight-Rate Problem of the United States*, House Document 264 (1937) chart 4 (John W. Alexander, *Economic Geography*, © 1963 Prentice-Hall Inc., Englewood Cliffs, N.J.).

the Eastern Territory that lower operating costs (per ton-mile) produced by the efficiencies of mass movement should be passed on to shippers in the form of lower rates. But this policy came under increasing attack from traffic agents in other parts of the nation arguing that the rate structure of Fig. 19.4 was discriminatory in favour of the Eastern Territory. In response to such pressure, spearheaded by the governors of southern states, the federal government in 1940 began gradually to reduce the differentials among the five regions. By 1952 all regions east of the Rockies were on equal footing.

2. The rate-group principle

One might suppose that along a railway radiating outward from a transportation centre, rates would gradually increase with distance. But this is not true. Instead, the places along each route are divided up into groups, all the points in one group having the same rate. Consequently, profiles of freight-rate structures follow stair-step pattern (Fig. 19.5), not a curve.

3. Class rates

The Interstate Commerce Commission authorises two types of rates: *class rates*, which apply to all items that move in small quantities, and *commodity rates*, which are applicable to a few items that move in large quantities. The rate-group principle is usually applied to class rates in such a manner that there is at least a loose correlation with distance. Consequently, class-rate structures tend to be somewhat symmetrical; this is illustrated by Fig. 19.6, which shows the pattern of class rates between Milwaukee and all points in Wisconsin. The lines on this map that connect points with equal freight rates are called *isophors*.

4. Commodity rates

Where the tonnage of a particular item is consistently large, the Commission permits a commodity rate that is lower than the class rate would be. For instance, in the rate structure within Wisconsin (Fig. 19.7) on loads of coal from Milwaukee (Appalachian coal that has come in by ship), the northern part of the state enjoys a more favourable position than does the western part – because of competition from railroads linking the northern area with other ports on Lakes Michigan and Superior.

Clearly, railway freight charges do not always increase at the same rate in all directions around a shipping terminal (isophors are not circular), nor do they increase consistently with distance in any one direction.

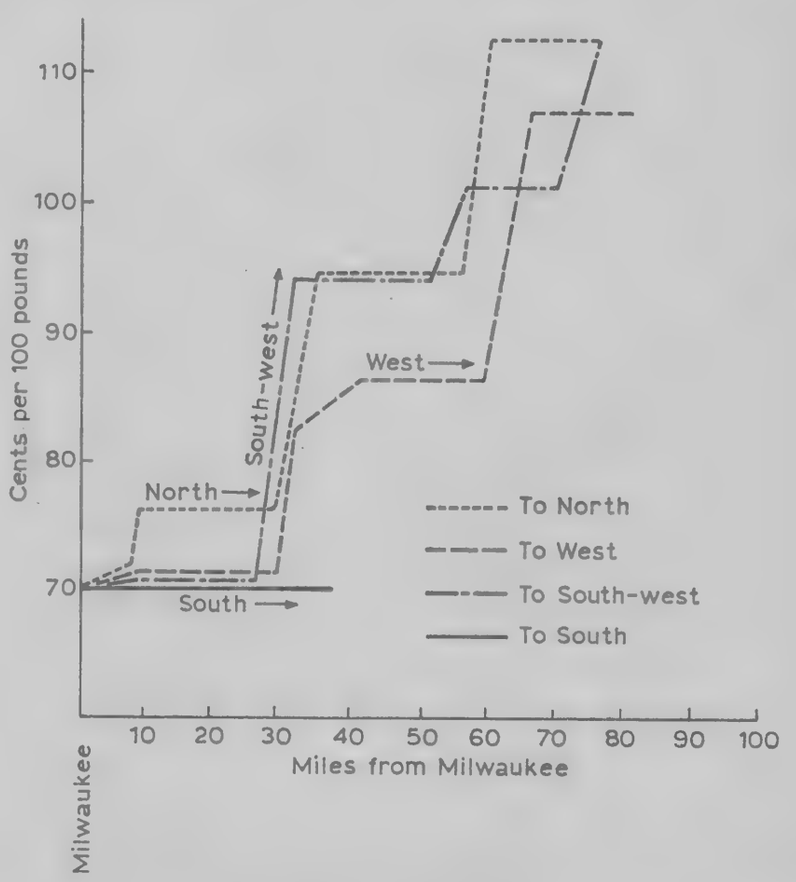


Fig. 19.5 Profile of railroad Class I freight rates, to/from Milwaukee, 1953 (*by courtesy of Economic Geography*)

5. In-transit rates

These are privileges granted by the I.C.C. to factories located between the origin and destination of a commodity. Sash and doors sold on the East Coast at one time were manufactured largely in Wisconsin factories from local pine. When Wisconsin's pine forests were cut away fifty years ago, these factories began buying

raw material from saw mills in Oregon and Washington. But this resulted in two separate rates, one from the Pacific Northwest to Wisconsin and another from Wisconsin to the East. *Since freight costs tend to increase with distance but at a decreasing rate,* the through charge on a 3000-mile trip is less than the combination of one fee on the first 2000 miles plus a second on the final 1000 miles. The 'in-transit' privilege permits a stop-over of raw material for processing before being shipped on to market – all at a *single through rate.* Such privileges are granted by the I.C.C. in only a few instances, but they spell the difference between life and death for factories located midway between raw materials and markets by enabling them to compete with factories located within the market areas. Most of Wisconsin's sash and door factories, for example, would have vanished with the pine forests but for in-transit rates.

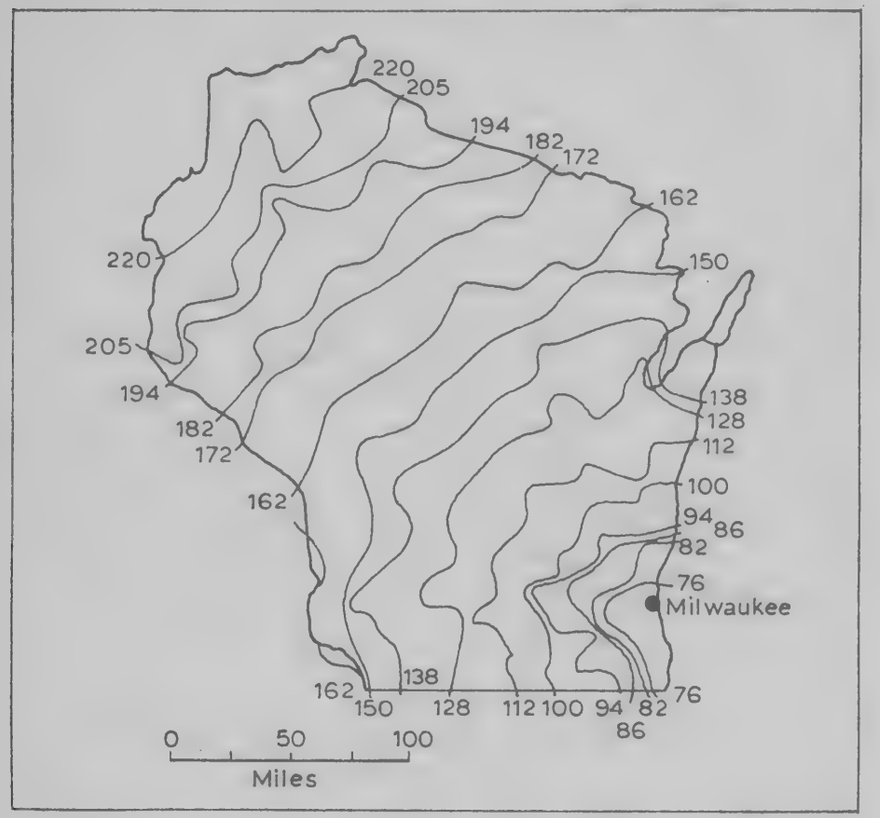


Fig. 19.6 Wisconsin freight-rate structure, 1953: railroad first-class rates to/from Milwaukee, carload rates in cents per 100 lb. (*by courtesy of Economic Geography*)

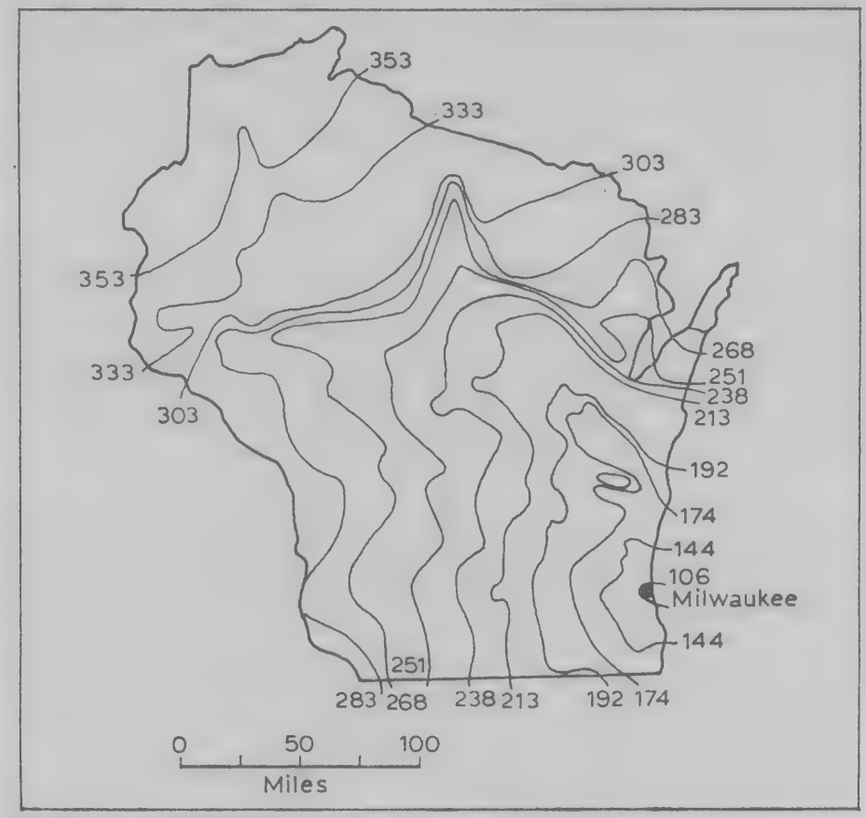


Fig. 19.7 Wisconsin freight-rate structure, 1953: coal rates to/from Milwaukee, carload rates in cents (*by courtesy of Economic Geography*)

Notes

1. Even if in practice zoning around cities is difficult to recognise today, it will be shown later that it is relevant to consider aggregations of such cities as single units about which agricultural production zones can be detected.

2. 'Freight Rates as a Geographical Factor in Illinois', *Economic Geography*, vol. 20 (1944) 25–30; 'Freight Rates: Selected Aspects of Uniform and Nodal Regions', *Economic Geography*, vol. 34 (1958) 1–18.

3. (Prentice-Hall, Englewood Cliffs, N.J.) pp. 473–6.

In reading Alexander's article one has to remember, as Grötewald said, that in von Thünen's time multiple freight rates depending on the commodity transported, etc., did not exist.

Another example where the price paid to ship a commodity is not commensurate with the mileage travelled is quoted by Michael Chisholm in 'Rural Settlement and Land Use'.[1] *He pointed out that the Milk Marketing Board pursues a price policy which tends to encourage production in areas remote from the main market. Indeed it does levy a charge on producers for transport and these charges do relate to collection from the farm and for delivery in the final market. But the charge varies according to which of sixteen regions the farm happens to be in. Further, the transport levy* does not reflect the true cost of transport *at all closely and makes no allowance for cost variations within the regions. Since the price paid per pint by the consumer is uniform throughout the United Kingdom, it is apparent that the Milk Marketing Board's policy is to encourage producers in remote areas away from markets who do not bear anything like the true costs of transportation of their commodity to the final market.*

In noting Chisholm's remarks, it should not be forgotten that modern transport and milk-processing technologies have enabled the Milk Marketing Board to pursue their deliberate policy of spreading production to remote areas of Britain. This contrasts strongly with conditions prevailing in the early nineteenth century, when milk, as a highly perishable commodity, had to be produced close to the market centre. This was the case as von Thünen set it out in his model. It was also the case in the area around London at that time, for the Rev. Henry Hunter, writing in 1811, was able to point to a series of concentric zones of agricultural production surrounding the city with milk production located in the innermost ring.[2] *Since the Rev. Hunter was describing a real-world situation, it is not surprising that, although many of the concepts von Thünen put forward in Part I of his book were seen to hold true in the formation of the rings around the city, some distortions were apparent. These, according to Hunter, were mainly explained by differences in soil conditions and in particular the presence of the flood-plain terrace gravels of the Thames and Lea, though there is no doubt that these rivers as relatively efficient means of transport (compared with horse and cart) also had their distortionary effect, as would be expected. But if soil conditions were important as an explanation of the distortion of the concentric zones, it did not mean that von Thünen had not appreciated this possibility. Though for the purposes of his model he first of all held that soil conditions were uniform on a flat plain (as shown in the left-hand diagram of Grotewald's Fig. 19.1), as a practical farmer he was of course aware of areal differences in the productivity of the land and in Part II of his book he showed that variations in soil fertility could cause production costs to vary substantially. Thus while soil conditions must be important in any explanation of land use patterns now or in the past, modern techniques of commercial farming, and in particular large-scale grain production, seem to imply the increased importance of terrain, a factor which von Thünen* always *held constant in his model. For example, modern cash-grain farming in the Midwest of the United States is completely mechanised from initial ploughing to final harvesting, and reasonably flat land must be present for efficient machine operation and for the maintenance of minimum rates of soil erosion. It is not surprising therefore that research workers have been prompted to ask just how important is terrain and in particular the flatness of the land in any explanation of the distribution of cash-grain farming compared with other factors, including soil fertility.*

J. J. Hidore has sought an answer for part of the United States Midwest in his paper 'The Relationship between Cash-Grain Farming and Landforms' (1963).[3]

The area selected for this study consists of eight states: North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri and Illinois. These states were selected for two reasons. First, within their boundaries occur portions of the largest contiguous concentrations of cash-grain production in the United States, and second, of the total acreage in the United States planted to corn, oats, wheat and barley, at least 45 per cent of each is within these eight states.

DEFINITION AND DISTRIBUTION OF CASH-GRAIN FARMS

A cash-grain farm is defined by the United States Census of Agriculture as being one on which at least 50 per cent of the value of all farm products sold off the farm comes from the sale of the specified grains.

Two measures of the distribution of cash-grain farms may be used. They are percentage of total farms classified as cash-grain and percentage of total farm area in cash-grain farms. The computed distributions are presented in Figs. 19.8 to 19.10.

The first map shows the distribution of the percentage of total farms classified as cash-grain on a county basis. The second distri-

bution is also percentage of total farms classified as cash-grain, but based upon state economic areas as the areal unit.[4] The third distribution is based on percentage of total farm area in cash-grain farms by state economic areas. The second and third distributions

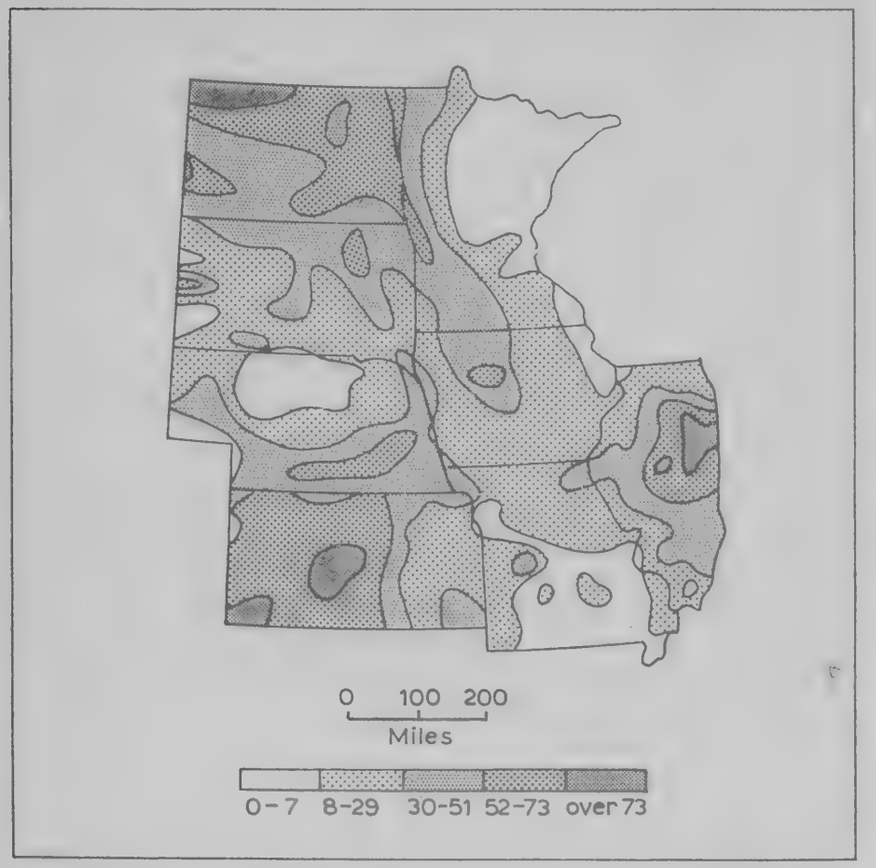


Fig. 19.8 Percentage of total farms classified as cash-grain, by counties

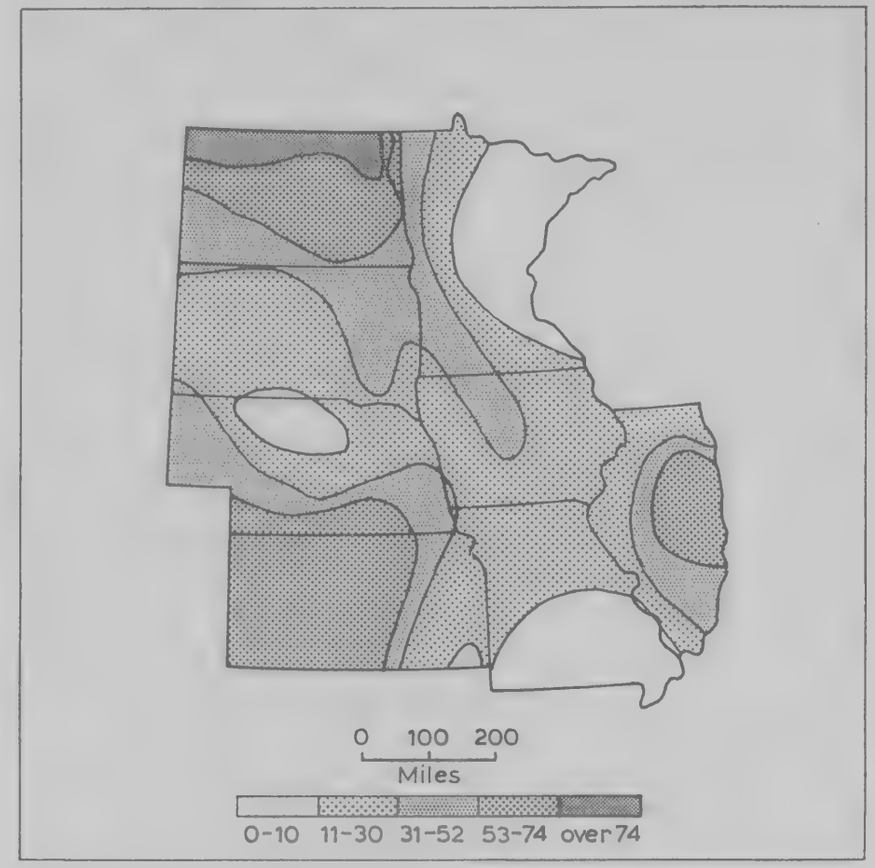


Fig. 19.9 Percentage of total farms classified as cash-grain, by state economic areas

are quite similar, having equal or nearly equal ranges, means, standard deviations and a coefficient of correlation between the two sets of data of +0·855. These similarities suggest that as the percentage of cash-grain farms varies, so also does the percentage of area in cash-grain farms.

The maps show three areas of relative concentration of the dependent variable. They are essentially east central Illinois, north-west North Dakota and western Kansas. Each of these areas is in part two standard deviations above the mean.

DEFINITION AND DISTRIBUTION OF FLAT LAND

With the very rapid increase in mechanisation of cash-grain farming, it appears as though the efficient limits of machine operation should be the most logical factor in determining the definition of flatness. However, this information is not easy to obtain.

Since data from farm equipment companies and related sources are not available, the definition of flat land had to be derived from existing literature, an investigation of which indicates a consensus that flat land may be acceptably defined as land with slopes of

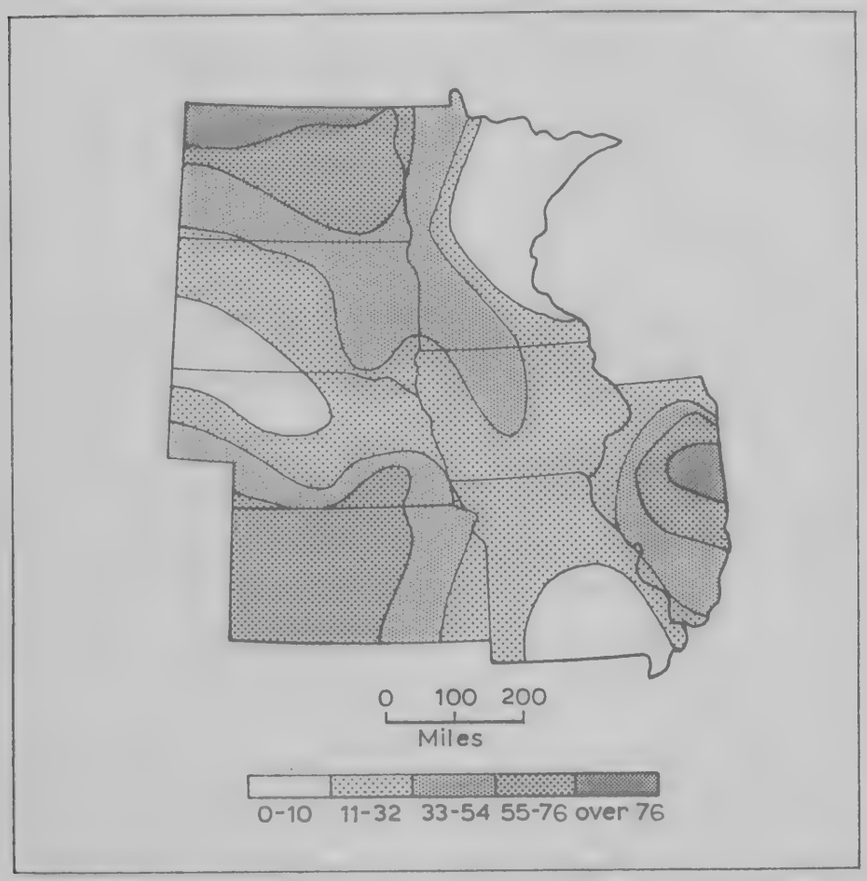


Fig. 19.10 Percentage of total farm area in cash-grain farms by state economic areas

three degrees (5 per cent) or less. In order to determine areas having slopes of three degrees or less, topographic maps were consulted.

The resultant distributions by county and economic area are shown in Figs. 19.11 and 19.12.

CORRELATIONS

After all percentages were derived for both flat land and cash-grain farms, a series of areally weighted product moment correlations were computed between the two distributions.

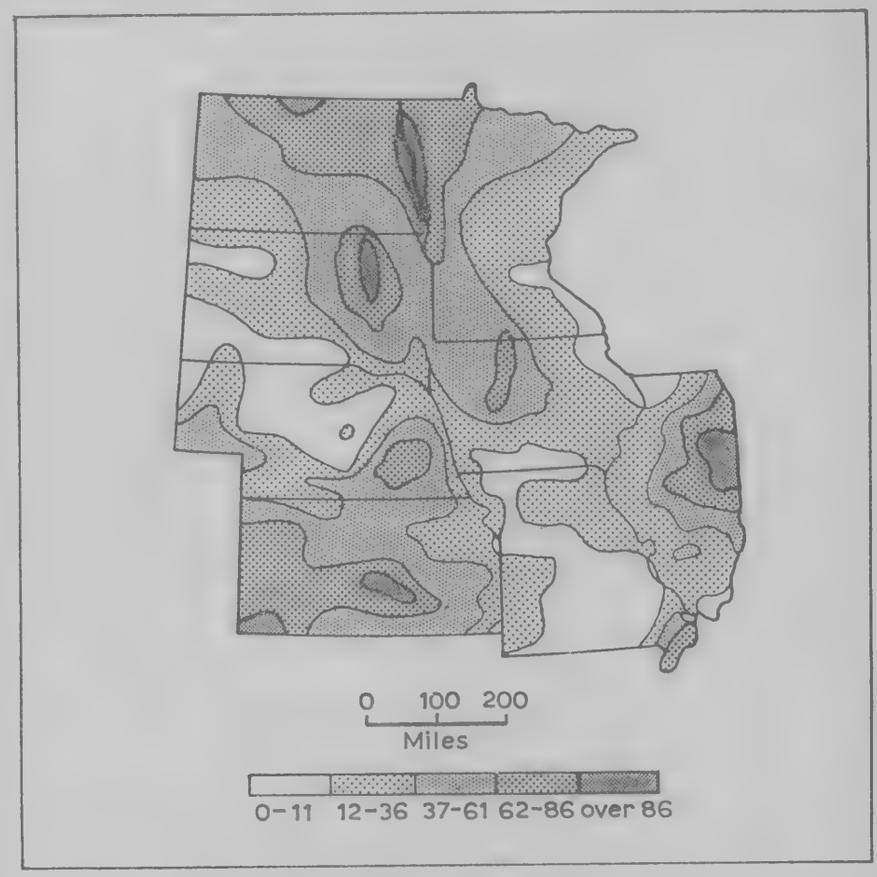


Fig. 19.11 Percentage of flat land per county

The correlations are based on two sizes of areal units, the county and the state economic area. The universe contains 730 counties and 72 economic areas. The coefficients of correlation (r) and determination (r^2) for the universe between percentage of flat land and percentage of total farms classified as cash-grain are:

(1) By counties $r = +0{\cdot}667$ and $r^2 = 0{\cdot}445$.
(2) By state economic areas $r = +0{\cdot}743$ and $r^2 = 0{\cdot}552$.

The value of the coefficients of determination indicate that the variation in flat land 'explains' about one-half of the variation in cash-grain farming.[5]

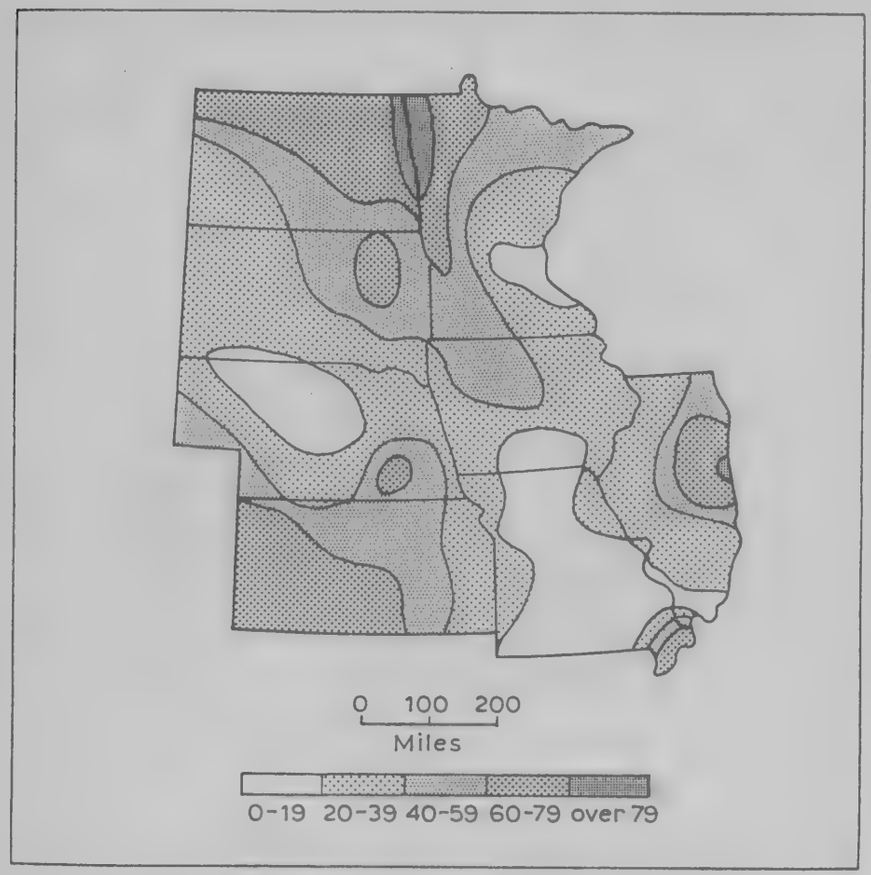


Fig. 19.12 Percentage of flat land by state economic areas

Tables 19.1 and 19.2 present the coefficients for individual states, based on county data and state economic area data respectively. The coefficients for South Dakota in Table 19.2 are the highest obtained for any state with either of the two dependent variables. In this case, the distribution of flat land 'explains' 95 per cent of the variation in the distribution of cash-grain farming in the state.

The second variable correlated with percentage of flat land is the percentage of total farm area in cash-grain farms. This was computed only at the scale of state economic areas as the data are not readily available on a county basis (Table 19.3). The coefficient of correlation for the universe between percentage of farm area in cash-grain farms and percentage of flat land is $r = +0{\cdot}746$ and $r^2 = 0{\cdot}561$. These two variables show the highest correlation for the universe as a whole. In this case the amount of flat land accounts for about 56 per cent of the variation in distribution of farm land in cash-grain farms.

The residuals map, obtained by plotting the error of the estimate for percentage of total farm area in cash-grain farms, suggests that the general tendency was to underestimate the amount of cash-grain farming in the areas of existing high density of cash-grain farms, and to overestimate in areas of lowest density of

Table 19.1

Value of Coefficients of Correlation and Determination between Percentage of Flat Land and Percentage of Cash-grain Farms for each State, Based on County Data

State	r	r^2
North Dakota	+ 0·971	0·943
South Dakota	+ 0·874	0·764
Illinois	+ 0·749	0·561
Kansas	+ 0·726	0·527
Iowa	+ 0·637	0·406
Nebraska	+ 0·585	0·342
Minnesota	+ 0·554	0·307
Missouri	+ 0·386	0·149

Table 19.2

Value of Coefficients of Correlation and Determination between Percentage of Flat Land and Percentage of Cash-grain Farms for States, Data by State Economic Areas

State	r	r^2
South Dakota	+ 0·974	0·949
Illinois	+ 0·926	0·857
North Dakota	+ 0·850	0·723
Iowa	+ 0·805	0·648
Kansas	+ 0·782	0·612
Minnesota	+ 0·730	0·532
Nebraska	+ 0·668	0·446
Missouri	+ 0·408	0·166

cash-grain farms (Fig. 19.13).[6] The residuals map also indicates where other independent variables may be more significant, as, for example, in north-eastern Minnesota and south-eastern Missouri. Both areas are over two standard errors from the mean.

TABLE 19.3

VALUE OF COEFFICIENTS OF CORRELATION AND DETERMINATION BETWEEN PERCENTAGE OF FARM AREA IN CASH-GRAIN FARMS AND PERCENTAGE OF FLAT LAND, DATA BY STATE ECONOMIC AREAS

State	*r*	r^2
South Dakota	+ 0·938	0·880
North Dakota	+ 0·894	0·799
Illinois	+ 0·860	0·740
Iowa	+ 0·771	0·594
Minnesota	+ 0·758	0·575
Nebraska	+ 0·712	0·507
Missouri	+ 0·682	0·465
Kansas	+ 0·681	0·464

In Minnesota the Laurentian Upland provides an area of poor soils and poor drainage which has practically no cash-grain farms. Flat land is present but the periods of glaciation have left much of this land covered by thin soils or virtually no soil at all.

The Mississippi flood-plain of south-eastern Missouri is another area with very poor association between flat land and cash-grain farming. The climate here favours a mid-latitude subtropical type of agriculture. Here, as in north-eastern Minnesota, is an area of flat land but with little grain produced.

Notes

1. (Hutchinson, 1962) p. 90.
2. For a fuller account, see G. B. G. Bull, 'Thomas Milne's Land Utilisation Map of the London Area in 1800', *Geographical Journal*, vol. 122 (1956) 28–9.
3. *Economic Geography*, vol. 39, pp. 84–9.
4. The state economic areas used are those established by the Bureau of Census. They are basically defined as 'single counties or groups of counties which have similar economic and social characteristics'.
5. The coefficient of correlation (r) is a dimensionless number which indicates the direction and degree of association between two variables. A value of zero indicates no association, and the more nearly the value approaches ± 1, the closer the relationship between the two variables. The + or − serves to denote whether the association is a direct or inverse relationship. The coefficient of determination (r^2) when multiplied by 100 indicates the percentage of variation in the distribution of the dependent variable (cash-grain farming) which may be attributed to the variation of the independent variable (flat land).
6. The error of the estimate is the difference between the actual percentage of area in cash-grain farms and a percentage estimated from the amount of that land in each economic area. Where the differences between the two values are greatest, the association between the two variables is poorest.

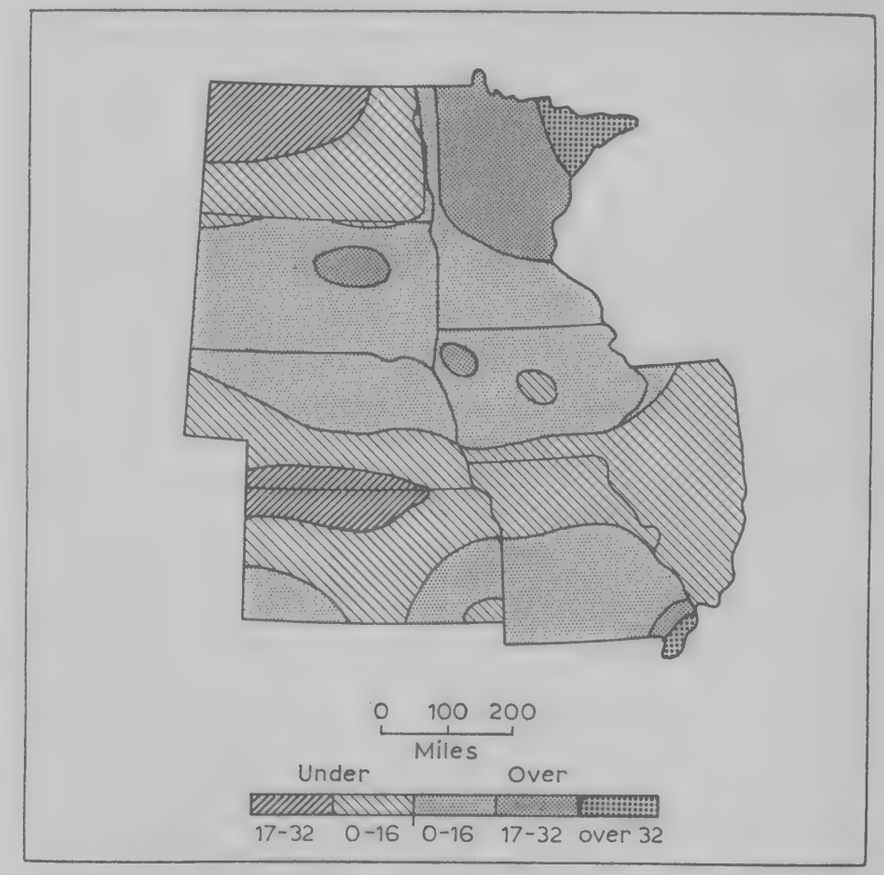


Fig. 19.13 Error of the estimate of percentage of total farm area in cash-grain farms

Thus while soils, drainage and climate are clearly factors helping spatially to determine the distribution of land in cash-grain farming, Hidore found that the distribution of flat land explained up to 56 per cent of the variation in the distribution of cash-grain farming for his study area. Von Thünen

of course could not foresee the importance of technological development on his propositions and thus the increasing importance of flat land, but he did appreciate the effect on land use of just those determinants Hidore considered relevant outside his 'flat land' explanation, including climate as well as soil. Climate, von Thünen realised, does affect the costs of plant and animal productions in various locations. It was only for the purposes of his initial discussion, however, that he assumed soil and climate to be constant, and his later acceptance of them as variables is not to negate entirely the value of his model. Furthermore, although von Thünen did not take account of it in the initial postulation of his model, he was also aware that it was necessary to recognise that farmers differed in their ability to manage the land just as they differed in the importance they attached to earning as high a profit as possible. He did, therefore, consider these factors in Part II of his book. For instance, he showed that farmers living near the margin of cultivation may accept a lower standard of living than those nearer the central city. This, he said, would depress the cost of production in the marginal areas and enable cultivation to be pushed farther into the remote areas than otherwise would be the case. Certainly in a real-world situation it cannot be assumed that the farmer is a totally rational economic man who has a single profit objective, omniscient powers of perception, reasoning and computation and the gift of totally accurate prediction. But while von Thünen recognised this, it is interesting that social and behavioural scientists have only recently come to give their attention to the importance of non-rational man.

In the field of geography the most important contribution to thinking in this field has been made by Julien Wolpert in his paper, 'The Decision Process in Spatial Context' (1964).[1] *The analysis of behavioural factors within the farmer where allowance has to be made for his ability to perceive the value of his environment as well as the fact that he will have multi-dimensional goals, was central to Wolpert's research, though as a secondary objective he did stress that the differences and variations in the attitudes of farmers are capable of coherent spatial analysis and are not randomly distributed among the population in the area used in his analysis, Middle Sweden. The theme of spatial dimension, though not of prime importance for us in considering this paper, does lie at the heart of the geographer's approach to his task. But Wolpert's central argument began as follows:*

The individual farmer, just as the industrial manager, must periodically make decisions with respect to the allocation of available resources among alternative uses. The farmer, functioning as a manager, must periodically decide how his land, labour and capital will be used, what will be the crop–livestock combinations, the investments in machinery and the other operating necessities. His goals with respect to income may vary anywhere from mere survival to optimisation. The information available to the farmer most likely does not include all of the relevant facts about costs and technology, and it cannot include knowledge of future events which will affect the consequences of his decisions. Within this environment of uncertainty, the farmer must make choices and must bear the burden of outcomes. All farmers are faced with similar problems, but actual decisions vary because farmers have different goals, different levels of knowledge and vary in their aversion to risk and uncertainty. These differences and variations have a spatial dimension and are not randomly distributed among the population.

THE SAMPLE POPULATION

Rapid changes are taking place in Sweden's agriculture and the focus of this dynamism is centred especially in the more urbanised and industrialised central portions of the country. Here, 10 per cent of the farms have been abandoned since 1956, and the survivors have been steadily moving in the direction of greater farming specialisation and intensification. Although the tempo of change is very rapid, the recent events have been documented with characteristic Swedish thoroughness, a factor of significance for both the normative and behavioural analyses.

The portion of Sweden included within Middle Sweden as defined for this investigation corresponds approximately to the zone known as Mellansverige, but has been confined to the full eight central counties (Fig. 19.14). The letters B, C, D, E, R, S, T and U are used to refer to the separate counties.

In this analysis, the areal surface of Middle Sweden's farming is the population. The purpose of the sample design is to survey the spatial properties of the decision behaviour of Middle Sweden's farming population. The sample population consists of a variety of farm situation types in terms of size, economy, site and settlement characteristics.

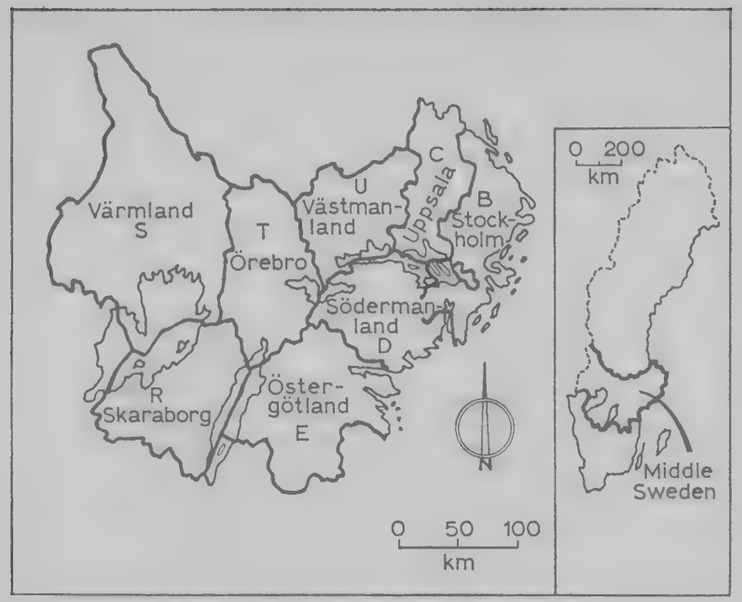


Fig. 19.14 The sample area, Middle Sweden, and its counties

THE NORMATIVE AND BEHAVIOURAL ANALYSIS

In a behavioural analysis, allowance is provided for the entire continuum of human responses from optimisation to that minimum of adaptation which may be essential for survival. Economic Man has a position at one extreme of the continuum. His position is predetermined in any situation in which the existence of an optimum may be demonstrated. The existence of a continuum provides an opportunity to explore the problem of determining the relative position of individuals on its scale.

The format of matched comparison serves as a suitable analytic framework for our discussion of the decision process. The decision behaviour of the population under study is compared with that of the matched control, Economic Man, who has the same supply of resources but is equipped with the necessary knowledge and foreknowledge to achieve his goal of optimisation. The factors of interest are the individual's goals or objectives and his level of knowledge which may be observed by means of a matching variable.

Performance levels in farming may be assessed according to a wide variety of criteria, only two of which were utilised in this analysis – organisation and technology. Farm organisation refers to the group of management decisions involving allocation of available resources among alternative activities. We shall define technology as the level of knowledge and application of crop and husbandry procedures. Finite farm resources of land, labour, capital and building facilities must somehow be combined to produce desired outputs at a given level of technology.[2] For each farm situation, with its finite set of resources, there exists, in addition to the actual organisation and technology followed by the individual farmer, an optimum counterpart which may be determined. Emphasis is reserved here for the gap between the organisation of resources on the actual farms, their technology and the optimum technologic level and organisation of the same resources, as measured by the differences in resulting labour productivity.[3] The productivity which the individual receives is matched with the optimum returns which the same resources would yield to Economic Man through optimum organisation and technology. The degree to which the departures are clustered spatially is a reflection of the systematic distribution of factors affecting decision behaviour. The focus is reserved here for these spatial attributes of the decision process.

Three spatial distributions are involved with respect to labour productivity, that resulting from: (1) the actual technology and organisation of farm resources; (2) the optimum or potential technology and organisation of the same resources; and (3) the gap between the actual and potential distributions. The optimum or potential productivity values were determined by means of a linear programming analysis for seventeen representative farm situations, and values interpolated for the remaining 533 farms of a systematic sub-sample (the J.E.U. sample), through regression estimation, see note 1, p. 347.

Linear programming, a mathematical technique permitting simultaneous consideration of many possible alternative organisational plans, specifies the most profitable plan consistent with available resources and the input–output relationships for the alternative enterprises. The range of possible enterprises included various crop and livestock combinations as well as forestry and afforestation of arable and pasture land, off-farm employment and off-farm investment. The linear programming procedure involves balancing the alternative enterprises according to their

relative contribution to net returns and their use of the finite supply of resources. The solution is in terms of the productivity or labour income which would result from the single most profitable combination of enterprises that is within the resource limitations, i.e. the potential productivity which the individual farmer could attain if his goal were optimisation and his knowledge perfect.

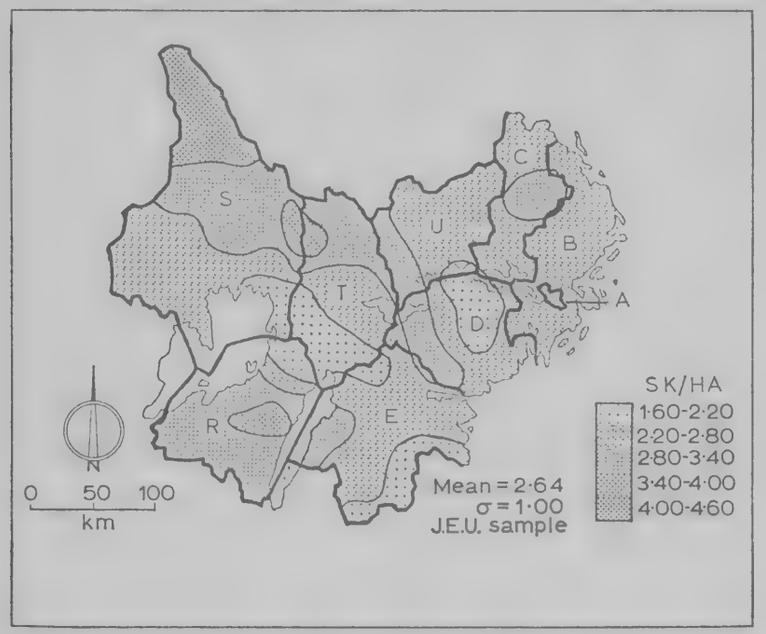


Fig. 19.15 Actual labour productivity: average per farm

The surface of actual productivity (Fig. 19.15) reflects the consequences of the outcome of the technology and organisation followed by the subsample farmers during the 1956–9 period.[4] Given their finite supply of resources, their individual goals and their levels of knowledge and foreknowledge, the actual productivity reflects the net return to their labour. Even in the zones of highest average productivity, as in R and northern S, the values fall considerably short of parity level with industrial workers' earnings.

The surface of potential productivity (Fig. 19.16) represents the theoretical limits of labour income with the present distribution of resources, i.e. the distribution of income if farm optimisation and technology were optimal. Any further increase in productivity would require additional resources. The major determinant of the potential levels is the ratio of the supply of capital to land and labour resources. Where the ratios are low, as in southern T, the poor structural balance limits labour productivity. The potential levels are highest in the northerly zone of Middle Sweden and in R, where 4.00 Swedish kroner (U.S. $0.80) an hour may be

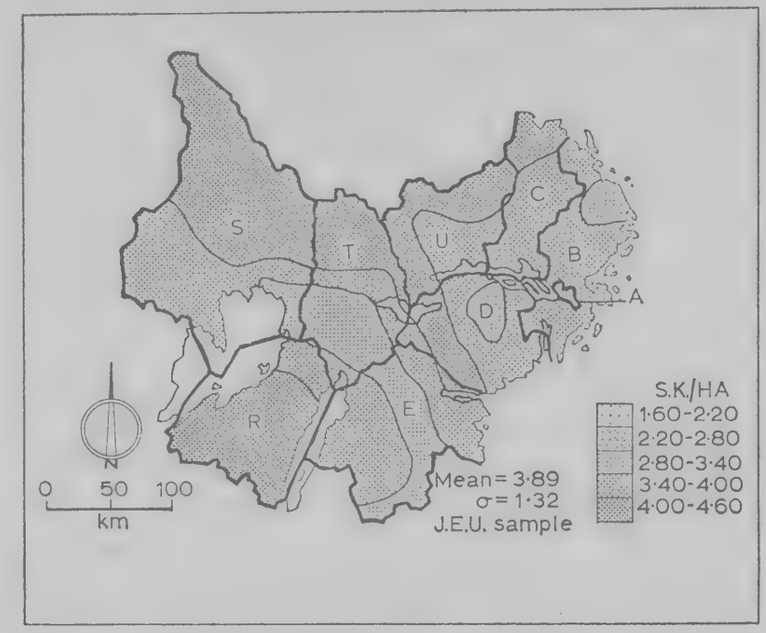


Fig. 19.16 Potential labour productivity: average per farm

earned within present resource limitations. The distribution of potential productivity is closely related to the structural variations in Middle Sweden's farms. Well-structured farms, in terms of the capital ratio, have higher potential levels because the need is not so great to organise resources so as to consume excess labour. When it is considered that the potential values indicated represent the optimum that may be achieved in the short run, it must be concluded that labour income on the sample farms is severely restricted, to an extent which varies spatially, by present structure and shortages of capital resources. This limit or ceiling appears especially noteworthy when one considers that nowhere in Middle Sweden do even these potential values approach parity with the general income level in Sweden. The variations in the distribution

are an indication of the varying degree of structural or resource maladjustment. Potential is merely a measure of available farm resources equated in terms of their income-producing possibilities. Clearly, income on Middle Sweden's farms is restricted by shortages of labour-conserving capital.

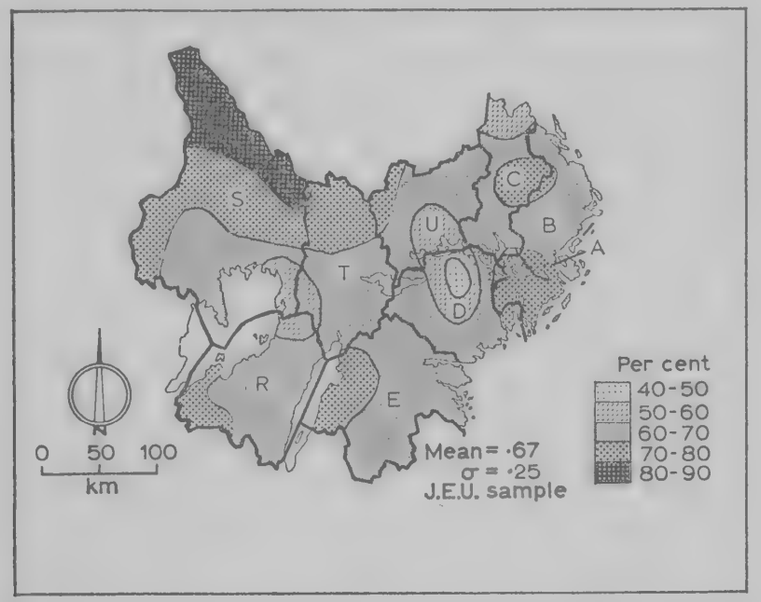


Fig. 19.17 Mean Productivity Index: ratio of actual to potential labour productivity

Imposed upon the structural maladjustment is the additional gap between the income that is realised and the potential income. We have combined the two distributions into a single ratio of actual to potential income (Fig. 19.17) to illustrate the extent and the variations in the gap from place to place. The sample population attains a lower productivity than their resoures permit. The surface, to which we shall refer as the Productivity Index or PI surface, represents essentially the human element in farming, the consequences of the decision behaviour of Middle Sweden's farmers with the resource differences removed. The gap as it varies spatially reflects the departure of the decision behaviour of our sample population from that of Economic Man and the degree to which the normative assumptions of optimisation and perfect knowledge must be relaxed in the behavioural analysis. The presence of variations in the PI surface leads one to suspect that there are significant spatial variations in the goals and knowledge levels of the population, and that the decision process has a spatial dimension.

For convenience, the PI surface may be stratified into reasonably homogeneous regions. An analysis of variance test of the distribution of PI values suggested the presence of five separate core areas with more or less distinct boundaries, suitable for a fivefold regional breakdown of the PI surface (Fig. 19.18). The regional means vary from 58·6 to 72·9 per cent (region 5 as opposed to region 1), but as one may infer from the standard deviations, the intra-regional differences between farms are considerable.

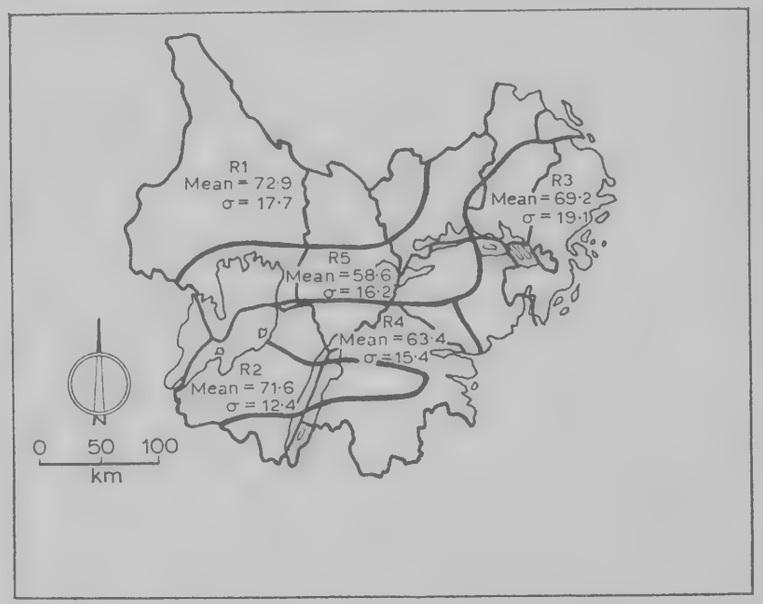


Fig. 19.18 Productivity Index regions. The mean regional productivity index values are indicated as well as the standard deviations

The two regions with the highest average values of PI (regions 1 and 2) are as different in terms of physical environment, farm structure and organisation and locational situation as is possible within Middle Sweden. Region 2 is an island of intensive farming and dense agricultural settlement. Region 1 is very sparsely settled with agriculture confined to areas where there occur

outcrops of non-morainal soils, such as silty clays and light sands in the narrow river valleys and lake shores. Similarly, with respect to the other regions, the rank order does not appear to be justifiable from the point of view of physical resources. These have been held largely constant or removed in the calculation of the PI values, so that the variations between the farmers themselves might be perceived and examined as a non-random but systematic distribution in the spatial dimension.

The five regions are delineated by the single criterion of PI value. This ratio, however, is a composite index reflecting nearness to or departure from economic rationality. Thus, although the regionalisation classifies areas according to a single dependent variable, we know that the classification extends to the explanatory variables also. In regions 1 and 2, for example, not only are the average PI values relatively higher, but because they are higher, then, of necessity, one would expect to find decision behaviour more closely approaching that of Economic Man. We would expect to find that resources are more knowledgeably combined and integrated so as to achieve higher profits. On the other hand, the lower average values of PI in regions 4 and 5 would be an indication that decision behaviour was less directed towards maximum profits, and that farm organisation and technology were hampered to a relatively greater extent by imperfect knowledge. Thus, the regions reflect a degree of internal homogeneity with respect to the dependent variable, the PI values, as well as the independent variables, knowledge and goals. The variance between regions reflects also differences with respect to both dependent and independent variables.

GOAL ORIENTATION

'Comparison now has to be drawn no longer to test the theory, but to test reality.'[5] The evidence presented would seem to suggest that reality is arational.[6] The hypothesis which was examined proposed that Middle Sweden's farmers do not achieve profit maximisation and that productivity is not limited only by the amount and combination of resources. Testing of the hypothesis for a sample population revealed that the average farmer achieved only two-thirds of the potential productivity which his resources would allow. With productivity not at the optimum level, then, it may be assumed that one or both of the prerequisites for economic rationality (perfect knowledge and optimising behaviour) are absent.

The alternative concepts of the 'optimiser' and the 'satisficer' are a central issue in this investigation in terms of their respective value as explanatory guides to the goal orientation of our sample population.[7] To a certain extent the optimiser concept has been introduced into economic geography by the tacit assumption that men organise themselves, their production and their consumption in space so as to maximise utility or revenue. Whenever the analyst projects his knowledge of the best or most efficient location of economic activities into an explanation of the actual distribution of phenomena and finds correspondence, he assumes, even though tacitly, that man is rational and that his objective reality corresponds with the subjective reality of the actor being observed.

Whereas the profit motive is present in every production decision in the sense that producers may be expected not to seek losses, maximisation may not be so universal.

The appeal of the satisficer concept or the principle of bounded rationality is clear. Based upon sound empirical investigations by social psychologists of actual economic behaviour, it requires nothing which is beyond the capacity of the human organism. The decision-maker merely classifies the various alternatives in his subjective environment as to their expected outcomes, whether satisfactory or unsatisfactory. If the elements of the set of satisfactory outcomes can be ranked, then the least satisfactory outcome of this set may be referred to as the level of aspiration adopted by the decision-maker for that problem. His search is complete and the action is taken. The theory suggests that aspiration levels tend to adjust to the attainable, to past achievement levels and to levels achieved by other individuals with whom he compares himself.

To some extent, the departures reflected in the PI surface may be attributable to the irrelevancy of the optimisation yardstick employed in its construction. There is greater justification for regarding the satisficer concept as more descriptively accurate of the goal orientation of Middle Sweden's farmers than the optimising criterion. The sample population is more concerned with alternatives which are good enough than in finding optimum solutions. The objective optimum solutions, as revealed by the linear programming analysis, are not perceived by the decision-

makers, who appear instead to have aspiration levels which they have a reasonable expectation of achieving, such as reasonable profits.

The satisfier model is not easily verifiable, for there is no simple way of determining the aspiration levels of individuals. Regional variations do exist, in the degree to which the farmers are commercially oriented, which are expressed by their lags in shifting with market changes. It may be argued, however, that the lag in response is attributable not to any lack of motivation but may be traced rather to imperfect knowledge which obviates the possibility of maximisation. It is debatable, therefore, as to whether we have an optimising population with imperfect knowledge or a population with goals other than pure profit maximisation. Certainly, the distinction matters little and to some extent is arbitrary, for goal orientation and level of knowledge are typically causally interrelated. Limited investigation of this argument revealed that non-economic considerations appear also to influence the decision behaviour of the sample population, especially with respect to work preferences.

KNOWLEDGE SITUATION

The assumption of omniscience implicit in the concept of Economic Man is not tenable with respect to the decision environment of Middle Sweden's farmers. Limited to finite ability to reason and compute, to perceive alternatives and to realise goals, optimisation may not be expected. The variations in the PI surface partially reflect these limitations as well as the effects of the uneven flow of information upon which the effectiveness of decisions is based. Information about prices, production methods and technological changes diffuses through spatial channels which discriminate between producers and create variations in knowledge situations. The emphasis given here to the communication process and the flow of information reflects the conviction that the individual's experience alone is insufficient to carry on productive agriculture. Productivity as a concept may be considered as an innovation diffused from group values.

Unlike the industrial firms who undertake the search for information on their own, the farm firm is usually somewhat isolated from the source of communicable information. The farmer is largely reliant for the information he needs upon the efficiency of a communication system which he is seldom able to control. The information comes to him largely without expense, in contrast to the situation of the industrial firm, but the lag on the transfer from expert to user is considerably greater, and perception and acceptance of the information are hardly automatic.

To a very great extent, information about recommended farm practices, new agricultural machinery, new seed varieties, expected costs and market prices originate for Middle Sweden's farmers among institutions in the Stockholm–Uppsala area. Although information is disseminated in a rapid and reasonably efficient manner, the communication process affects farmers unequally and this unevenness is significant spatially.

The diffusion process proceeds by steps, from the experts in the core zone of Stockholm–Uppsala to the disseminating organisations' central offices, then to the local county offices which are situated typically in the county's leading city. This initial process takes place rather rapidly without significant differences from one county to another in Middle Sweden. The transmission then proceeds in stepwise fashion, directed at first to the larger farmers and those situated within the major agricultural districts of each county. An intra-county diffusion process is set in motion which involves filtering and imitation through the farmers' ranks so that considerable lag intervenes until small farmers who are isolated are made aware of new information. Farm size, membership in the farmers' organisations and location with respect to other farmers all apparently influence the network of communication over the surface. The uneven spatial distribution of farmers according to farm size, membership and proximity to other farmers is proposed as the primary determinants of the uneven flow of information and the variations in knowledge levels. These governing rules for the communication process may be incorporated within a simulation model in order to estimate stochastically the spatial variations in the supply of information at a given time.

The demand for information may be considered as well as the spatial supply process by assuming, for example, that in the areas of highest potential productivity (Fig. 19.16) the innovation of rationality would have started earliest, developed fastest and be least restricted by a ceiling. The communication stream would be fastest in the areas of highest potential productivity (responding

to the greater demand for information) and the spatial lag would be greatest where the relative advantages of shifting in the direction of rationality were least (where there is less demand for information). The distribution of opportunities, as revealed by the surface of potential productivity, is suggested as the major determinant governing the demand for information.

It is not sufficient, therefore, merely to substitute the behavioural concepts of finite ability to perceive and absorb information for the untenable assumption of omniscience in characterising the knowledge situation of Middle Sweden's farmers. Spatial biases exist in the communication channels which enable farmers in some areas to receive information more rapidly and thus to have access to larger funds of knowledge upon which to base their decisions.

THE ENVIRONMENT OF UNCERTAINTY

In the discussion of communication channels, it had been assumed that knowledge was essentially available to be disseminated and that knowledge levels were dependent upon the intensity of diffusion. The decision-maker is limited further, however, to *ex ante* information and must bear the consequences for the gap in his knowledge about future events.

In addition to the objective uncertainty which occurs because of unpredictable change and variations, notice must be taken also of the uncertainty apparent to individual decision-makers which may be considerably different. The uncertainty environment to which the decision-maker reacts is dependent not only upon the inherent instability of phenomena but upon the state of knowledge concerning this instability. Uncertainty has a spatial dimension, therefore, not only because the inherent stability of phenomena may differ from place to place, but also because communication channels diffuse information unevenly in space, and because perception varies.

The farmers of Middle Sweden must make decisions in an uncertain environment. These decisions appear with respect to crop and livestock combinations, production techniques and other practices which affect farm income and survival. The consequences of their decisions during the 1956 to 1959 period are reflected in their productivity distribution for those years (Fig. 19.15). Most likely, with *ex post* information, their decision behaviour would have been significantly different.

Five types of potential sources of objective uncertainty may be identified which are relevant for Middle Sweden's farmers. Uncertainty may be introduced through instability in (1) personal factors, e.g. farmers' health and ability to work; (2) institutional arrangements, e.g. government policy, landlord–tenant relationships; (3) technological changes; (4) market structure; (5) physical factors, e.g. weather, blight and other environmental variables. These potential sources of uncertainty have been reduced systematically, however, through social insurance, legal tenant contracts and long-term government price policies in the agricultural sector. Only yield variability arising from weather uncertainty remains as the most significant unknown in the planning environment of Middle Sweden's farmers, and its impact is spatially differentiated. Climatological records for the past are an insufficient tool for predicting weather accurately enough to ensure optimal selection from among cropping alternatives. Wide and irregular fluctuations of output occur in Middle Sweden's farming which cannot be controlled or predicted. Resources are committed and expended at the beginning of a crop season when output is uncertain and will not be known for months. The period under investigation, 1956 to 1960, was, in aggregate, quite normal for Middle Sweden. Some farmers had ideal weather and good yields, but for many others the cost of uncertainty was wasted effort, expenditures and unstable income.

Coefficients of yield variations $\left(\frac{\sigma}{M} \cdot 100\right)$ have been determined for feed grains (barley, oats and mixed grains) and winter bread grains (wheat and rye), to illustrate their respective degree of variability during the period 1921 to 1960 (Figs 19.19 and 19.20). The fact that little correlation exists between the yield variability of the different crops may be used to advantage by farmers in deciding upon suitable crop combination and rotational systems. Reducing uncertainty through diversification and complementarity are strategies that are extensively followed by Middle Sweden's farmers. In addition, the correspondence is generally very high between yield variability and crop emphasis. In the areas of relatively higher risk for feed grains, emphasis is shifted to crops subject to comparatively less local variability, as hay.

The separate coefficients of variation for the crop groups have been combined into one measure and weighted by prices (1960), average yields and the acreage devoted to the individual crops (Figs 19.21 and 19.22). The average percentage departure in crop revenue that may be expected each year may be interpreted from Fig. 19.21. In the high-risk zone in the north-east coastal

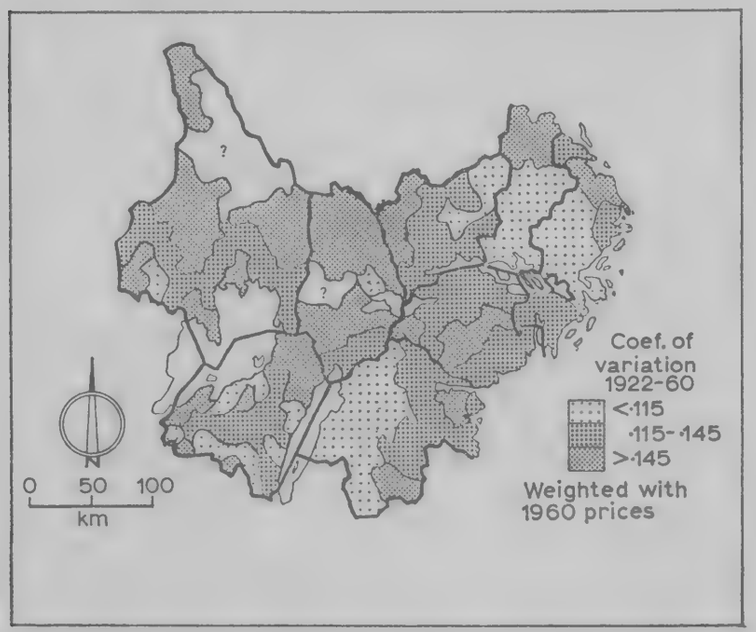


Fig. 19.19 The yield variability of feed grains. (Data drawn from *Permanent skordekadeskydel* [Permanent Protection against Crop Damage], Statens offentliga utredningar, 1958: 5, Ministry of Agriculture, Stockholm.)

area, for example, total crop revenue may be expected each year to depart by ± 25 per cent from the average. Immediately to the north, the average variation is only ± 15 per cent. The difference may be traced to several causes. The areas differ little in terms of variability for individual crops, but in the high-risk area a greater proportion of the acreage is devoted to higher-risk crops and average yields are higher. The risk is relatively lower in the north-west because yields are normally low for most crops and much of the arable land is planted in lower-risk hay. If greater emphasis were given to winter grains, a much higher degree of variability could be expected.

A relatively similar distribution pattern results when the expected variability is translated into a measure of average revenue departure per hectare of arable land (Fig. 19.22). In the central plains area of E, farmers may expect a normal variation of ± 150 Swedish kroner per hectare (U.S. $12.00 per acre) each year. An

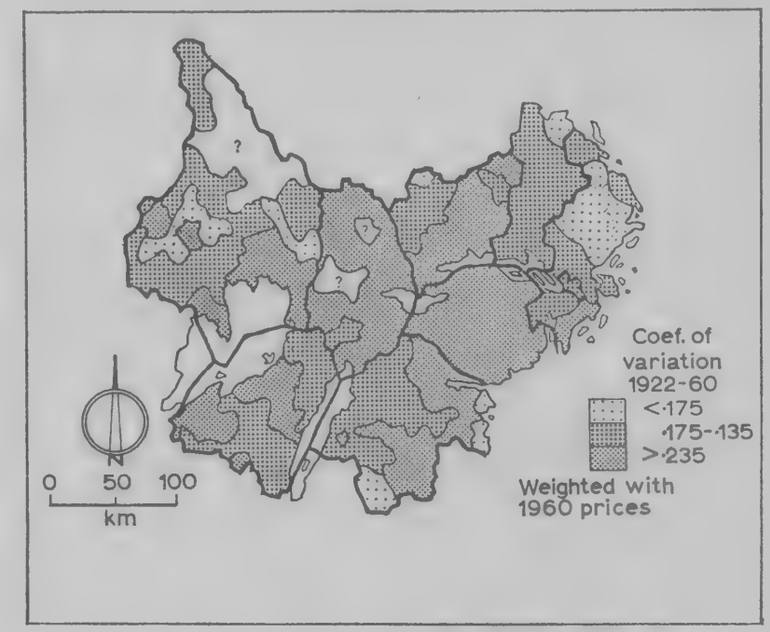


Fig. 19.20 The yield variability of winter bread grains. Source: as Fig. 19.19

average farm there with 30 hectares of arable land (74 acres) should anticipate that revenue from crops will vary an average of U.S. $900 yearly from the long-term average unless broad diversification is followed. The revenue per hectare varies less in most other areas of Middle Sweden, for less emphasis is given to wheat, oleiferous plants and other risky crops. The typical farmers in E appear to prefer cultivating the more lucrative crops despite the risk. Greater stability in income is apparently considered more desirable by those in central S.

The uncertainty resulting from unpredictable weather variations is an extremely significant factor in Middle Sweden's

agriculture. Even with the present pattern of diversification, farmers must expect their revenue from crops to vary between 15 and 25 per cent from the average. The factor that appears to account to the greatest extent for the spatial variations in this distribution is the emphasis given to individual crops and combinations in the rotation plan, the environment of uncertainty which the individual decision-makers determine for themselves.

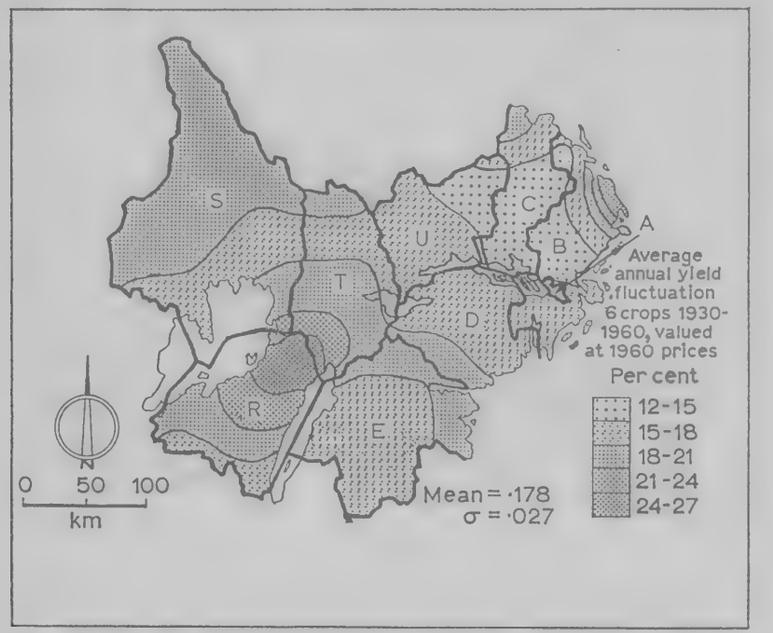


Fig. 19.21 The variability of crop yields. Source: as Fig. 19.19

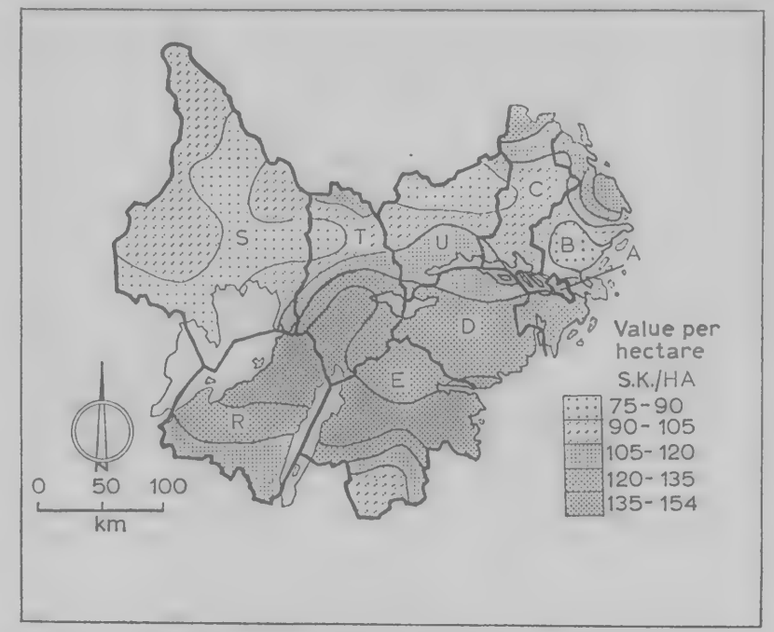


Fig. 19.22 The variability of crop yields: deducted value per hectare. Source: as Fig. 19.19

Spatial inequalities exist at any given time in the extent of foreknowledge possessed by the farmers of Middle Sweden, owing in part to the inequalities in information diffusion. Foresight is more easily achieved in areas subject to less unpredictable change, but knowledge of the methods which are available to reduce its amplitude plays a significant role as well. The communication channels which disseminate information bring news about the technical and economic measures which may lead to greater control over the unknowns. Therefore, we may expect to find that, in the areas of more intense interaction and communication, farmers are better acquainted with the risks involved with alternative enterprises and the means by which uncertainty may be reduced to acceptable risk levels.

The presence of uncertainty eliminates the possibility of profit maximisation. Even if the intention is to optimise, imperfect knowledge prevents its realisation. The existence of risk and uncertainty about the consequences of alternative courses of action makes it necessary for the decision-maker to consider not only the desirability of outcomes but also their probability of occurrence. Courses of action must be determined, the outcomes of which are both sufficiently certain and desirable to satisfy the decision-maker. The quest for income stability is apparently at least as important to Middle Sweden's farmers as the pure profit motive. We are aware that in a behavioural investigation the assumption of perfect knowledge must give way to a knowledge continuum. Similarly, recognising that individuals vary in their aversion to risk and uncertainty, the mechanical optimiser must now yield to a utility continuum.

Attitudes towards risk and uncertainty are not easily observable.

Theoretically it should be possible to analyse the cropping and other decisions made by Middle Sweden's farmers and determine their respective positions on the continuum scale. This type of evidence is revealing but hardly sufficient alone. Instead, we shall attempt an indirect procedure by analysing factors which are hypothesised as causally related to position along the continuum.

Research by social psychologists and others has revealed that certain personal and situational characteristics are related to individuals' attitudes towards risk aversion. Age, for example, has been found to be a particularly important variable. It has been consistently confirmed in these investigations that older people tend to a greater extent to select courses of action which involve lower degrees of risk. In Middle Sweden, a selective migration has all but emptied certain rural areas of young farmers. Therefore, one would expect, if position on the utility continuum were dependent purely upon the age distribution of farmers, that a very definite regional hierarchy on the continuum scale could be expected. This is the sense in which the existence of spatial variations in goal structure is implied. Aspiration levels are difficult to measure and assess, but their major dimensions may be perceived through knowledge of factors which affect their formulation.

Factors other than age play decisive roles in the satisficing scale and several have spatial distributions which are non-random. Equity position influences the ability to take chances and still survive. Size of family may be expected to influence attitudes towards stability and security. Farmers who are located in areas subject to periodically destructive weather conditions, such as early frost, might tend to select those alternatives which minimise risk. Other factors undoubtedly affect the type of compromise which farmers select, such as amount of available working capital, farm scale, opportunities for efficient diversification or shifting, the time preference in consumption, institutional restrictions in the capital and marketing systems, the time sequence of poor years in the past and personal attitudes towards conservatism or gambling. Some farmers require a certain income to ensure the survival of their farm, and they accept the risks necessary to achieve this level. For others who may be considering a change of occupation, nothing less than income parity with non-farm workers is sufficient. Thus, although income goals themselves may not be revealed through simple measurement, some aspects of their distribution among Middle Sweden's farmers may be revealed through analysis of the distribution of some of the factors mentioned above.

Various alternatives are available to farmers who wish to temper their desire for profit with a measure of security. When uncertainty may be reduced to definite risk probability, maximisation at least of expected income is possible. Techniques have been determined for calculating maximum incomes at various risk levels. Even when absolute uncertainty is present and goals are defined in terms of maximum–minimum incomes or other security levels, it is still possible to specify procedures for selecting the best means for realising these goals.

For our purposes, the satisficer type of approach appears more appropriate. The farmers of Middle Sweden do not necessarily select the best course of action for realising their goals with respect to profit and security, but follow a process of adaptation based upon the feedback of information. It is proposed that the farmers arrive at some normal production programme for their farm enterprises based upon their subjective knowledge of alternative courses of action with respect to technology and selection of enterprises, their resource position, their aspiration levels and the pressures and restrictions imposed by institutions. This normal pattern is maintained as long as returns remain within a range which is satisfactory to the farmer. When a crisis occurs, the farmer is jarred out of the previous pattern and shifts in a direction so that, once again, a satisfactory outcome may be expected. The normal production programme pattern need not be a static one, for normal growth and expansion may be anticipated and readjustment might require long-term change.

The established programmes include some provision for stability of income which may be inferred from the distribution of crop rotation patterns and crop and livestock combinations. There are various means which farmers have at their disposal to achieve stability despite the uncertainty of their environment. The creation of credit institutions, crop insurance and marketing agreements have had an important impact, but our major concern is with the measures used by individual decision-makers.

Farmers, as other decision-makers, may postpone the making of decisions until further information is available. Some of the farmers delay entering into new lines of production until prices appear to be stabilised over a period of time. Many decisions in

farming cannot be postponed beyond a very limited date, such as planting or harvesting dates. The retailer and manufacturer may rely upon the constant feedback of information and may make relatively short-run decisions. Farmers must commit large proportions of their resources at one time and have little opportunity of shifting crops, for example, when more weather information becomes available.

The more common means used to avert or diminish risk and uncertainty involve diversification, shifting and flexibility. Maintaining a cash or semi-liquid reserve with which to meet sudden emergencies allows for a degree of flexibility but only at the expense of profits in the short run. Shifting may involve transference from a risky crop to one which gives more stable yields, or the farmer may leave farming entirely and shift to a more certain occupation. Through diversification, a combination of enterprises may be selected which reduces the risks that may be associated with specialisation. Diversification may succeed in reducing the total variability in returns to a level which is less than the variability of the separate enterprises.

Specialisation, even in dairying, is accompanied by considerable risk, and in hay production the risk is much higher, but when the two enterprises are combined, the total income variation is considerably less than either alone. Similarly, the lucrative production of rape seed may be undertaken at a comparatively low risk when combined, for example, with wheat and potato cultivation. By means of diversification, farmers may balance potential profits against potential risk to find a suitable combination of enterprises. Diversification, as with respect to the other measures, may bring about a greater level of stability but at the expense of efficiency or profit. Uncertainty and risk have a cost which detracts from opportunities.

The departures reflected in the PI surface, as well as the spatial variations, may be partly understood through knowledge of the uncertainty environment to which Middle Sweden's farmers react, and through the means they have created to provide a controlled environment of relative stability.

Notes

1. *Annals of the Association of American Geographers*, vol. 54, pp. 537–58.
2. Capital, as employed in this study, refers to working capital, assets in liquid or semi-liquid form (which can be translated into liquid assets within one crop season), e.g. cash on hand, value of livestock and crops planted. The working capital values are calculated by subtracting from total assets the real-estate value of farm property, buildings and long-term inventory items such as machinery and tools. Therefore, working capital measures the assets available to the farmer for direct use in enterprise selection.
3. Labour productivity refers to net labour returns to the farm family per equivalent man-hour after a deduction of 5 per cent interest return on investment. Optimum technical levels were determined from estimates of potential regional yield levels for crops and milk as reported in *Proceedings of the 1960 Agricultural Investigation* (Uppsala: Department of Agricultural Economics, Royal Agricultural College, 3 March 1962). To achieve the potential yields, the farmer would find it necessary to fertilise according to a plan based upon a soil survey of his arable land, and selective breeding and proper feed mixture for dairy cattle would be necessary as well.
4. The four-year period, 1956–9, rather typical in terms of weather and market fluctuations, was used as the base for both the optimum and actual productivity calculations.
5. August Lösch, *The Economics of Location* (New Haven: Yale University Press, 1959) p. 363.
6. The term arational is used to refer to situations outside the sphere in which rational distinctions or judgements may apply, as distinguished from irrational.
7. H. A. Simon, *Models of Man* (New York, 1957) pp. 196–200.

Thus, as Wolpert has demonstrated in his study, Middle Sweden's farmers did not achieve profit maximisation, nor were their aims solely directed to such an objective. Perfect knowledge was denied to them by the existence of unpredictable change and a lag in communication and perception of information. The decision behaviour of the farmer reflected not only the objective alternatives which were available, but his awareness of these alternatives and the consequences of their outcomes, his degree of aversion to risk and uncertainty and his system of values. As Wolpert concluded, 'the concept of the spatial satisficer appears more descriptively accurate of the behavioural pattern of the sample population (of farmers) than . . . the concept of Economic Man.' (See note 1 above.)

These considerations of the behavioural aspects of the farmer, along with the other real-world variables mentioned above, do not then appear in the von Thünen model as it was initially set out in Part I of his book, though, as has been pointed out, they were later taken into consideration by him with the realisation on his part that, applied singly or together, they would profoundly modify the pattern of concentric land use zones. At the same time their application would not invalidate the basic principles of the original model.

REAL-WORLD VARIABLES DEMONSTRATED IN THE MODEL

In a recent publication, 'The Spatial Expansion of Commercial Agriculture' (1969),[1] *J. R. Peet considered that these variable factors exerting diverging pulls on the concentric zones as set out in the original model, would cause them to all but disappear. However, he suggested that 'while transport costs continue to form a major part of the total costs of producing and marketing crops, at least some semblance of a concentric zonal system remains'. On the basis of this assumption he then theoretically analysed the influence of some of the variables already discussed above on the von Thünen zones of the original model. He began by suggesting that it can be assumed that the influence of the physical environment acts primarily through spatial variations in costs and revenues, and that acceptance of this notion means that the effects of the physical environment are easily assimilated into the basic von Thünen location theory. He continued:*

With a homogeneous environment the rent of each crop decreases regularly outwards; the rent of crop *C* (segment II, Fig. 19.23) decreases from 600 at the market to zero at the 'agland[2] boundary', and its 'iso-rent' lines[3] from exact circles around the market. Environmental variations, by influencing costs and revenues, distort this regular decrease with distance. Imagine, for example, that the yield of crop *C* rises around point *X* in such a way as to cause its iso-rent lines to form circles around that point, pushing the rent *C* yields above the level of rent given by crops *A* and *B* (which do not change as point *X* is approached). Within a 'critical' rent level on the hill of high rent for *C* about point *X*, at which the rent given by *C* equals that of the surrounding crops, the production of *C* is substituted for production of *A* and *B*. In a transport-dominated cost environment, the physical factors thus cause local deviations from regional distance-determined norms. As transport costs decline in relative importance, the size and frequency of these deviations are increased.

Dropping the assumptions of a single market and equal transport costs in all directions causes easily predictable variations in the simple von Thünen-type model. Subsidiary towns (segment III, Fig. 19.23) have their own micro-zonal systems, although some micro-zones may be absent due to competition from nearby great zones (oriented to the dominant city) in which economies of scale are obtained. Lower transport costs in any direction, due, for example, to lower cost media, or faster agencies in the case of perishable commodities, cause the zones to move outwards along the lines of lower spatial resistance (see examples in segment V, Fig. 19.23).

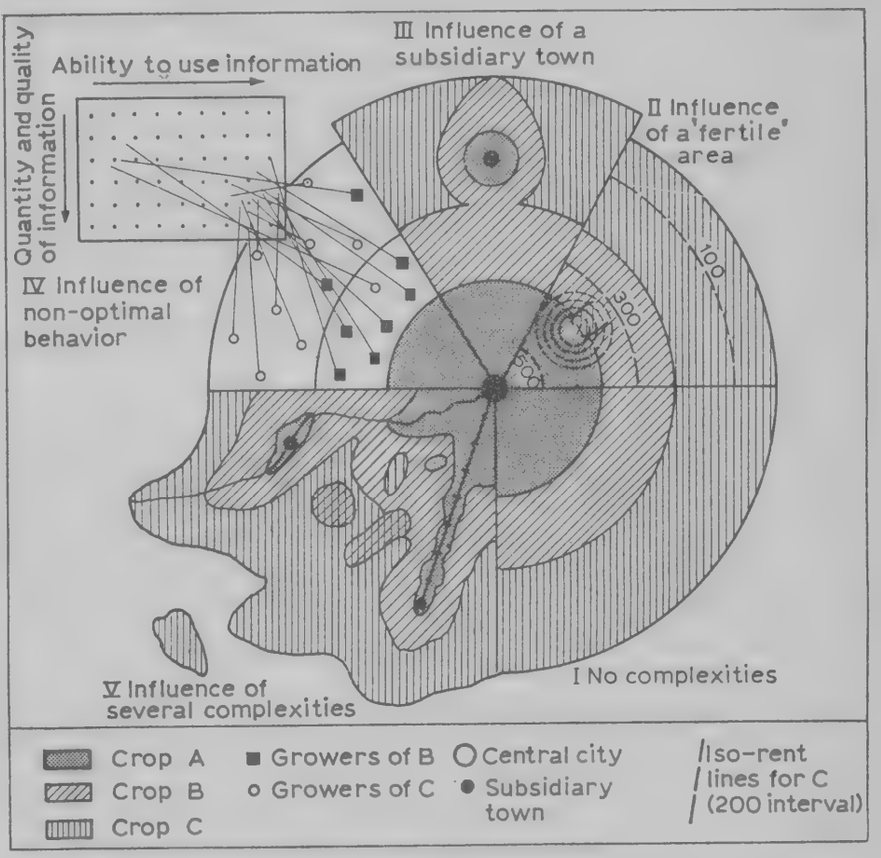


Fig. 19.23 The von Thünen zones under several complicating conditions. (Behavioural matrix after Allan Pred, 'Behaviour and Location: Foundations for a Geographic and Dynamic Location Theory', pt I, *Lund Studies in Geography*, series B, *Human Geography*, no. 27 (1967) 25, 76.

Variations in farmers' entrepreneurial abilities, and in the amount of knowledge, are here thought of primarily as boundary-blurring forces. The behaviour-types of farmers in such transition zones may be visualised with the use of Pred's 'behavioural matrix' (Fig. 19.23, segment IV). The majority of farmers in the *B* and *C* crop zones (the only zones considered – to keep the diagram

simple) produce crops *B* and *C*, and occupy positions in the lower right-hand corner of the matrix, signifying a high level of knowledge and ability. However, some farmers near the margin produce the crop of the contiguous zone, rather than the optimal crop, and survive due to the fact that at the point of juncture between two crops both yield *some* profit. Enclaves of the sub-optimal production may, therefore, be due to knowledge and ability variations.

An attempt to portray the effects of all the disturbing influences on the simple von Thünen zones of segment I is made in segment V (Fig. 19.23). The extent to which many complicating factors, operating at the same time, alter the appearance of the concentric zones, will depend on their cohesiveness and on their power relative to the power of the opposing forces which act to put a cost on all deviations from norms set by position relative to the market.

Peet then considered the role of geographical change in the modified zonal system he had described, which could take the form of changes in crop sequences or contractions or expansion in one or a number of zones. (This might in turn cause contractions and expansions in the whole agland.) He found it convenient to divide the forces of change into those acting through alterations in demand and those acting through supply and to assume that (in the first instance) only one force change occurs at any time.

A rise in demand for any crop, or group of crops, will lead in the short run to a rise in price. This will have two main effects on supply: (1) it will be profitable to intensify production within the zones already raising the crop; (2) both higher prices and the resulting intensification raise the rent yielded by the crop, and this creates pressure for zonal expansion. The contiguous zones 'resist' such expansion and contraction in their supplies through price and rent rises. Eventually, a new equilibrium is reached in which the zone producing the newly-in-favour crop is usually wider, and the total agland larger.

It might be noted that agland expansion can take three general forms: (1) land of lower fertility (or land more difficult to cultivate under prevailing technology), *within* the outer frontier, may be broken into cultivation – hence the clearing of the woodland in medieval and early-industrial Europe; (2) virgin land at the frontier may be joined to the regional, national or international economy – thus the movement of the international frontier into the temperate continental interiors under the pressure of rising industrial demands in nineteenth-century Europe and North America; (3) a high-rent agland may capture part of a lower-rent agland (land only lightly used because of poor communications with one market may become efficiently linked with a larger alternative market and begin producing in greater quantities for this market).

Changes in the conditions of supply may be divided into production-technological and transport-technological change. Dunn has argued that production-technological changes that are yield-increasing, tend (obviously) to restrict the production area, while technological improvements that lower either cost of production or transport tend to extend the production area. Transport-rate reductions do this by raising local prices, thus making production at the frontier profitable (rent-yielding).

Until now, changes in the conditions of demand and changes in the conditions of supply have been assumed simply to occur, as though in a causal vacuum. Yet changes in crop demand, and changes in the technological conditions of supply, are intricately linked. Only one main chain of relationships will be followed through here – that which might be postulated to result from an increase in crop demands under conditions of vacant land (or land held by less powerful cultures) beyond the existing spatial limit to farming profitability. This case most closely resembles that facing the urban-industrial countries in the nineteenth century.

A rise in crop demands and a temporary rise in prices makes more intensive production possible on lands formerly farmed in a relatively extensive manner, and also causes inner zones, frequently producing large bulks per acre, to move outwards. These changes increase the volume of interregional trade between large zones specialising on those crops which distance from the market suggests and the central urban-industrial demand nucleus. A rise in trade volumes, besides increasing economies of scale in transport, also may justify innovations thought previously to have entailed too high risks. Such economies and successful innovations reduce costs to transport companies. Because of the tendency for monopoly in this industry, these cost reductions may not be immediately passed on, but eventually duplication of services makes rate declines more probable, and declining rates consolidate or even, under certain conditions, extend the outward zonal advances

made. Indeed, the combination of rising demands, which support prices, and a transport system which is becoming more and more efficient, is essential for protracted expansion.

On the one hand, if demand were to rise but ton–mile costs did not fall, the increasing distance over which goods would have to be moved would add to market prices. Eventually, food prices would become so high that city growth would be choked off. The traditional method of avoiding this was the localisation of cities on water, which provides a line of lower spatial resistance to movement and allows an elongated agland to develop.

On the other hand, declining transport costs in a period of stagnant demand would eventually overbalance supply, causing prices to fall, thereby removing much of the inducement for expansion. This stifling effect may, however, set in slowly. Depending on the slope of the curve expressing the relation between increasing capital and labour inputs and increasing crop outputs, a small decline in price may make much of the intensive crop production of the inner zones unprofitable. In such a situation agland growth may merely replace the previously intensive supply system and not significantly add to the volume of production.

Under conditions of rising demand *and* falling transport costs the von Thünen model forecasts rapid outward movements of the supply zones about a market. The exact nature of these movements depends upon the relative strengths of the forces causing changes in demand and those leading to changes in supply. Nevertheless, it can be said that both transport innovations which affect particular crop zones (which expand and 'shunt out' neighbouring zones) and general transport innovations particularly benefit agricultural production far from the market. In this way the agricultural frontier is pushed into vacant or lightly-held distant lands.

THE MODEL RELATED TO THE CONTEMPORARY WORLD SITUATION

As a logical outcome of the above argument, J. R. Peet concluded by suggesting a world urban-industrial nucleus occupying much of Western Europe and north-eastern North America which emerged in the wake of the Industrial Revolution during the latter part of the nineteenth century. He stated that this may be hypothesised to have a similar role in the international economy to that played by the single city in the von Thünen model. The urban-industrial nucleus exported manufactured goods and was the dominant market for food and raw materials produced on commercial farms in the surrounding countries. Commercial agricultural specialisation in these countries can therefore be expected to have been determined largely by economic distance from the dominant market.

Michael Chisholm, in 'Rural Settlement and Land Use' (1962),[4] *quoted from an article by J. T. Schlebecker*[5] *which argued along lines similar to those of Peet. Schlebecker, who was concerned to provide a theoretical framework for the study of agricultural history, showed that von Thünen's ideas do have a real relevance. First, he drew attention to the fact that in past centuries there were a large number of urban concentrations of small size, each with its own zones of agricultural production. In the modern world there are far fewer, but each is considerably larger and influencing a great part in the world. Like Peet he looked upon Western Europe and the eastern United States as one metropolitan area in the world context. Secondly, Schlebecker, in a notion which is closely related to the first, noticed the sequence of land uses which may be experienced by any one area and observed the way in which they repeat the order of intensity of farming types. As far as America was concerned, he observed that in the nineteenth century 'on a single farm a man might in his own lifetime be in cattle country, wheat country, corn country and dairy country'.*[6]

It must be apparent, as Chisholm has rightly pointed out, that this concept of 'the moving frontier and the attendant changes in land use can only be understood in terms of location principles which provide a necessary framework for analysis'.[7]

Notes

1. *Economic Geography*, vol. 45, pp. 283–301.
2. The agricultural hinterland of the city which supplies it with produce.
3. Lines of equal rent.
4. p. 189.
5. 'The World Metropolis and the History of American Agriculture', *Journal of Economic History*, vol. 20 (1960).
6. Ibid., p. 203.
7. *Rural Settlement and Land Use*, pp. 189–90.

As with rural land use patterns, the location of industry is a spatial problem and therefore capable of analysis by the geographer. Again as in the case of the last chapter, the geographer can usefully employ theoretical analysis through the examination of a basic model. This enables the fundamental proposition suggested in the model to be properly examined, stripped of confusing detail. Once this has been understood, theory can be married to the complexities of the real-world situation.

THE LEAST-COST LOCATION MODEL

The classic model used for tackling the problem of industrial location is that of Alfred Weber. He put this forward in his book 'Über den Standort der Industrien' published in 1909.[1] Central to an understanding of his ideas is the notion that industry will locate itself at the point of least cost. The model is set in an isolated state where the natural resources to be processed conform to a von Thünen ring system around a number of given market centres. This naturally suggests the existence of spatial variations in economic advantage with regard to both supply and demand, as was seen in examining the basic model in Chapter 19.

Weber, for the purposes of his model, divided the basic natural resources the manufacturer has at his disposal into two kinds. In one group he placed materials like mineral fuels and metallic ores. These he called 'sporadic' materials on the supposition that their availability would occur only at a limited number of sites. In the other he placed what he termed 'ubiquitous' materials such as water, sands, clays, etc., which would be very much more widely available. Weber considered that these natural resources have either 'gross' weight if they lose weight in the manufacturing process, or 'pure' weight if their weight is not diminished by processing. From these considerations he was able to draw a distinction between more localised and more dispersed materials and between 'less mobile' and 'more mobile' materials.

Weber then said that inside his heterogeneous environment (cf. von Thünen's isolated state), manufacturers would locate industries at the point of least cost and that they would do this in response to three location factors. These factors were:

(a) *Transport.*
(b) *Labour.*
(c) *Agglomeration or deglomeration.*
(Agglomeration means the drawing together of industries of like kind for the purpose of gaining mutual benefits. Deglomeration refers to the reverse process.)

Weber also assumed, like von Thünen, that transport costs increased directly with distance and that the weight of the material carried would be part of that cost factor – thus his concept of ton-miles. The point of least transport cost therefore would be that at which the combined weight movement involved in assembling materials from their sources and in distributing the products to the markets was at its smallest.[2] To indicate whether the point of minimum movement was closer to the source of the materials used in manufacture or closer to the market, Weber devised what he called a 'material index'. He calculated this by dividing the weight of the localised or sporadic material inputs by the weight of the final products. If the resulting index was greater than 1 then the point of movement minimisation would be closer to the source of materials; if it was less than 1, then it would be closer to the market.

If an industry had a high ratio of labour costs to the combined weights of material inputs and product outputs (this was what Weber termed his labour coefficient), then labour would tend to attract the industry to a location other than the movement minimisation one, provided the savings which resulted in labour costs per item of production exceeded the additional transport costs which would result in marketing. Weber also added that manufacturers might also deviate from the minimum transport and labour cost locations and develop their industries at a third minimum point if economies in costs could be achieved by agglomeration or deglomeration. Weber considered that agglomeration might lead to economies because manufacturers grouped in close proximity could create a common pool of skilled labour, indulge in common purchasing and selling policies, and generally co-operate in other ways. On the other hand, the grouping of industries in urban zones could lead to pressure on space, rising land prices, and eventually encourage deglomeration.

On pp. 61–6 of Weber's book the location possibilities (or orientation) of a number of hypothetical types of industry are considered, bearing in mind weight movements, labour coefficient and agglomeration economies.

F. E. I. Hamilton[3] *has summarised this section of Weber's work in the form of a table:*

TABLE 20.1

INPUTS, LOCATION FACTORS, AND LEAST-COST LOCATIONS FOR WEBER'S HYPOTHETICAL INDUSTRIES

	Inputs			*Other factors*			*Orientation*					
		Sporadic										
Industry	*Ubiquitous materials*	*'Pure' materials*	*'Gross' materials*	*Labour*	*Agglomeration*	*Deglomeration*	*Materials*	*Intermediate*	*Market*	*Labour*	*Agglomeration*	*Deglomeration*
A	1								*			
B	2+								*			
C		1					?	?	?			
D	1+	1							*			
E			1				*					
F		2							*			
G			2=					*				
H			2≠+				*					
I	1+		1				?	?	?			
J	2+											
K		unspecified		*						*		
L	,,	,,	,,		*						*	
M	,,	,,	,,		*						*	
N	,,	,,	,,			*						*

The symbols used are:
+ for 'or more' materials.
= for equal weight losses for gross materials.
≠ for unequal weight losses for gross materials.
* for definite locations.
? for alternative (equally viable) locations.

The above table is then explained as follows:

This clearly reveals the main features of Weber's model. Industries A, B and D, that process ubiquitous materials, are oriented to the market since any other location would involve the unnecessary costs of transporting ubiquitous materials in the form of a bulkier product. Weber argues that industry F, using two localised 'pure' materials, will locate at the market for the same reason. Where differentially weight-losing materials form the only inputs, then entrepreneurs will locate their industries near the sources of materials, ideally, that lose most weight (industries E and H). Where two or several gross materials lose equal weight (industry G), entrepreneurs will locate at some intermediate assembly point. The location optima, however, are not so obvious either for industry C which manufactures a 'pure' sporadic material incurring (*à la* Weber) equal transport costs wherever the industry is located, or for industries I and J which process both localised weight-losing and ubiquitous materials. In this last case the 'minimum movement' location will depend upon the input ratio of gross materials (and their weight-losses) to ubiquitous materials. Industry K, with a high labour coefficient, is located in the cheapest labour location, industries L and M are linked and are attracted to a common centre on account of agglomeration economies, and industry N seeks out an 'out-of-town' location because of large land requirements.

Notes

1. Translated by C. J. Friedrich as *Alfred Weber's Theory of the Location of Industries* (Chicago, 1929).
2. This is why Weber's theory of industrial location is sometimes called a minimum movement or minimum energy model.
3. In *Models of Geography*, ed. R. J. Chorley and P. Haggett (Methuen, 1967) p. 371.

While Hamilton's above summary deals essentially with hypothetical industries, Weber's location model has also been explained by the use of empirical data drawn from a specific industry, that of zinc smelting. In his 'Location Analysis in Human Geography' (1965), Peter Haggett[1] has written:

The central argument in Weber's theory of industrial location was that sites would be selected *ceteris paribus* so as to minimise unnecessary movement, i.e. that sites would represent minimum-energy positions. We may conceive movement as being made up of three separate components, the *distance* to be moved, the *weight* of the material inputs or outputs to be moved, and the *effort* or cost of moving given materials over unit distance. We may combine the first two components, distance and weight, to give a *gross movement input* which we can measure conventionally in *ton-miles*; and we may combine this index with effort to give a *net movement input* which we can measure in dollar-ton-miles. For convenience this final index of movement will be referred to as Q.

Table 20.2 shows the calculation of net movement inputs for our case industry, and we can see how first tonnage and then freight costs modify the amount of movement over standard distances for the four major inputs and major output of a smelter. The most striking contrast is between heating coal, which, when hauled over the same standard distance as fireclay, represents a net movement input some twenty-four times greater, a contrast explicable in terms of its greater importance in the industrial process and its higher freight rates. Clearly this is an oversimplification: in practice, freight rates tend to be convex and stepped with distance (see above, pp. 293–6), but the net movement input provides a rough guide to the relative movement that we are trying to minimise.

TABLE 20.2
DERIVATION OF NET DISTANCE INPUTS*

Parameters:	*Location*	*Distance (miles)*	*Weight (tons)*	*Freight rate (dollars/ton-mile)*	*Net distance inputs, Q*
Outputs:					
Slab zinc	e	1·00	0·54	2·10	1·14
Inputs:					
Zinc concentrate	α	1·00	1·00	1·00	1·00
Reduction coal	β	1·00	0·37	1·10	0·41
Heating coal	γ	1·00	1·08	1·10	1·19
Fireclay	δ	1·00	0·10	0·50	0·05
Total inputs		—	2·55	—	2·65

* Example of zinc-smelting industry.

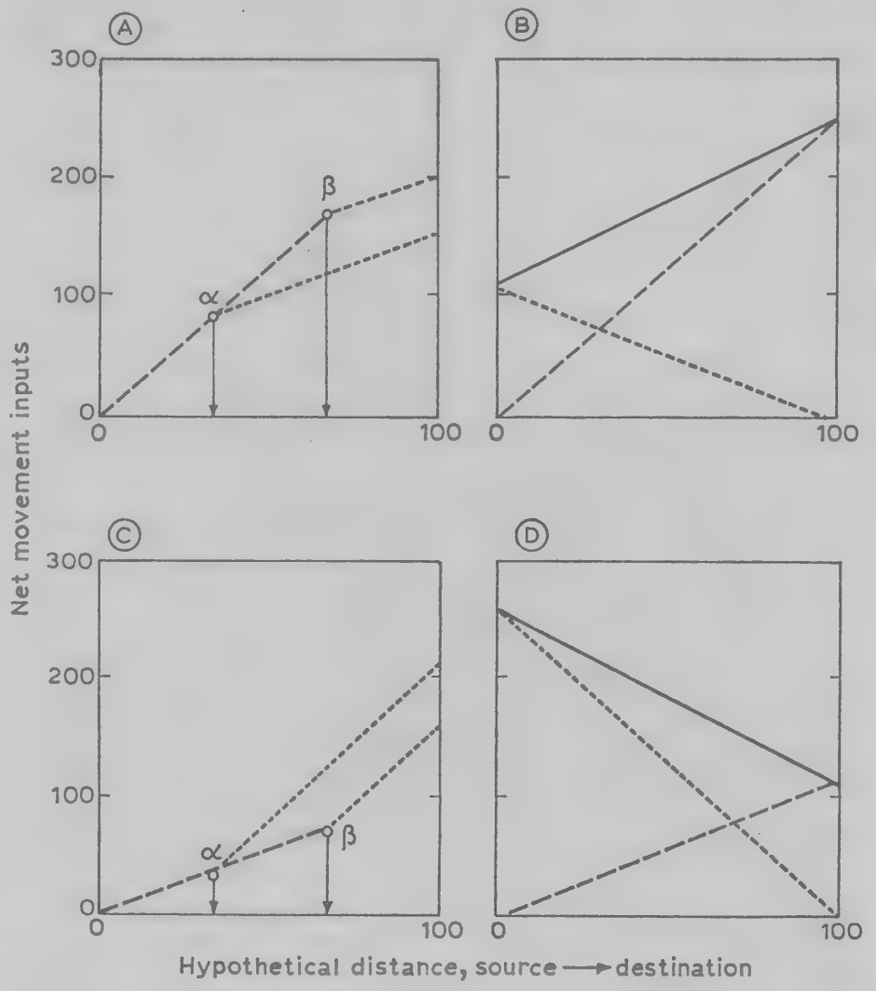


Fig. 20.1 Weberian weight-loss analysis of market or destination orientation (*A*, *B*) and material or origin orientation (*C*, *D*)

MOVEMENT MINIMISATION IN A TWO-POINT CASE

The concept of net movement inputs may be used very simply in a two-point case by relating the movements towards the plant or assembly movements, Q_a, to the movements away from the plant after processing of finished products or distribution movements, Q_b, to give an orientation index (V). Here V is given as $V = Q_a/Q_b$. When the value of V is greater than 1·0 the orientation of the plant is towards its material sources, and when less than 1·0 to its product destinations.

In our empirical case (Table 20.2) the sum of the four assembly movements (2·65) and the sum of the single distribution movement (1·14) give an orientation index of 2·32, suggesting that the basic orientation of the industry is towards its sources. The position may be shown very clearly by plotting values of the net movement inputs for the industry over a hypothetical distance of 100 miles. In Fig. 20.1A it will be seen that the total movement increases as the location is shifted from location α near the material source to location β near the destination. The general relation of the net movement index is shown in Fig. 20.1B where the solid line represents the sum of both inward assembly movements and outward distribution movements. This shows clearly that location at the source is in this case the minimum energy position.

Were the ratios exactly reversed, i.e. the orientation index V is 0·43, then the locational position would be likewise reversed. Fig. 20.1C and Fig. 20.1D show the solutions for this reversed situation with in this case location β being more logical and the most effective minimal location being at the destination.

Note

1. pp. 143–5.

Haggett's analysis is then another way of demonstrating Weber's least-cost or minimum energy model. At its heart lies the orientation index, or as Weber called it, the material index; but Haggett only considered this in relation to one case using empirical data and it is valid to ask how efficient such a locationary index would be if applied to other industrial examples taken from the real world. This question is partially answered by W. Smith, who took sixty-five industries in Britain and compared the results of the application of the index to them with location reality. He summarised his work in his paper 'The Location of Industry' (1955),[1] and early on pointed out that without doubt, in the case of some industries there exists a correlation between the index and location:

The weight of the raw sugar extracted in a sugar beet factory is one-eighth of the weight of the materials (beet, coal and lime) used in the extractive process; the weight of the butter, cheese and manufactured milk products is in the aggregate one-sixth of the

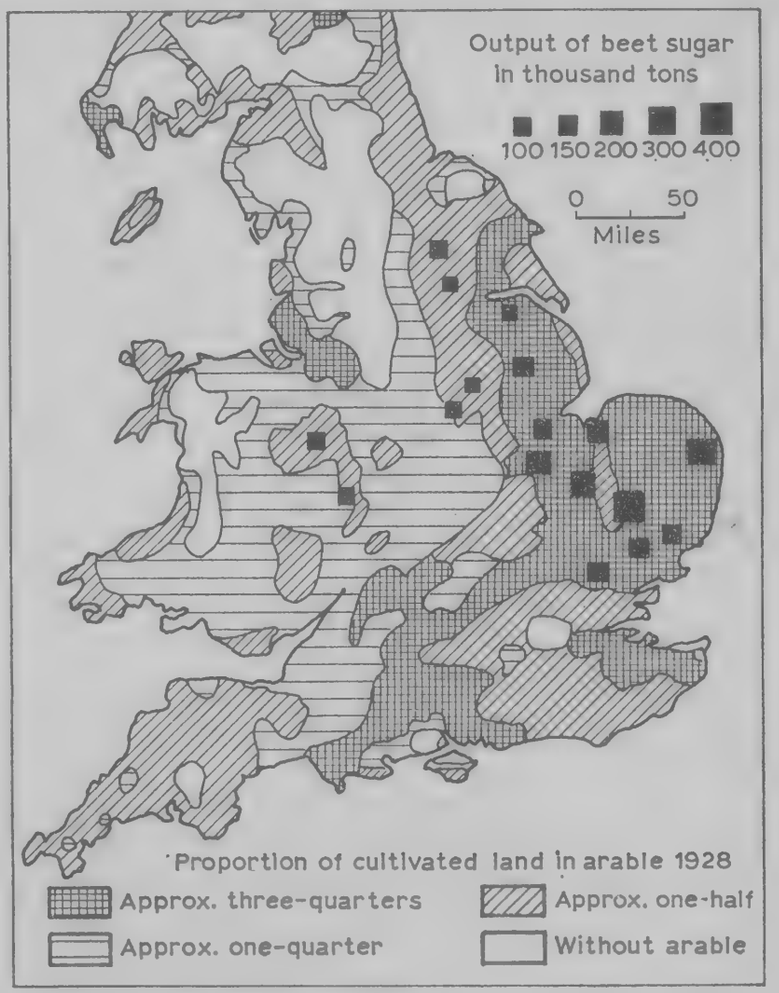


Fig. 20.2 Distribution of sugar beet factories in relation to arable land

weight of materials used in the milk factory; the weight of pig iron is between one-fourth and one-third of the blast furnace charge. In Weber's terminology they have a material index of 8, 6 and 3–4 respectively. The watershed which Weber adopts between industries orientated to materials and industries orientated to markets is a material index of one, that is, in which weight of product equals weight of materials. The sugar beet factories are in the midst of the beet-growing districts, though a few are peripheral (Fig. 20.2); the blast furnaces are on ore-field or coalfield, with the single exception of the Ford plant at Dagenham; the milk-processing factories are in the midst of dairying districts, with the exception of those urban plants manufacturing waste milk surplus to urban consumption (Fig. 20.3). Each of these industries displays a large loss of weight of the main product during manufacture and each is located close to supplies of raw material. The few plants located elsewhere are capable of rational, if not always of economic, explanation. Two points arise immediately out of these examples. The first is that they can all be described as primary industries, the initial handling of a material direct from mining or from farming. It is in this initial processing of the material that large losses of weight arise. I will return to this point shortly. The second is that these large losses of weight occur in respect of the main product and that in arriving at them by-products and waste have been ignored. In each of these industries there are by-products of some importance: beet pulp for stock feeding from sugar extraction, basic slag for fertilisers from iron smelting, whey and butter-milk for stock feeding from cheese and butter making. These by-products are bulky, are very much lower in value than the main product, and for precisely these reasons are most extensively used within a relatively short distance of the processing plant. They do not invalidate the conclusion that such plants must be located close to their raw materials.

There are many other industries whose raw materials lose in weight during manufacture but not to this extent. The loss in grain milling varies with the extraction rate which itself varies with the kind of flour it is intended to produce: for white flour the loss is 30 per cent, not including coal. In fish curing it is 31 per cent for herrings and more for other fish; in mineral oil refining 22 per cent. These are also primary industries in the sense that they handle materials for the first time, though not necessarily at the site of primary production. It is not an accident that they are port industries processing materials imported by sea, though there is some milling and tanning from home resources and in inland districts. The weight loss from these imported materials is very much less than in beet sugar, pig iron and milk product manufacture which handle only (or largely) home resources. If loss were as high in this second group as in the first, then the initial process-

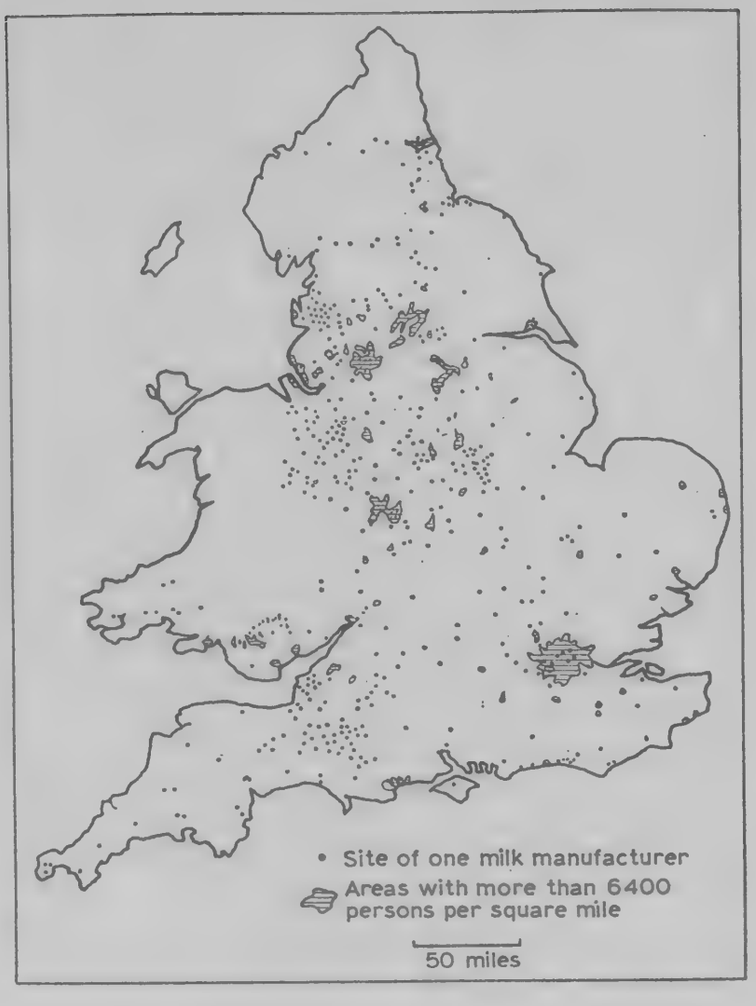


Fig. 20.3 Milk manufacturers, 1934–5

ing would be at the site of the raw material abroad, as it is, for example, in cane sugar, in which loss of weight during the extraction of the sugar juice is of the same order of magnitude as in sugar beet. The extraction of sugar from the sugar cane, therefore, is an industry of the sugar plantations and only the refining of raw cane sugar is a British industry. It is significant that the loss of weight in leather tanning is 34 per cent from British hides coming fresh from slaughterhouses with fat and hair still attached, but only 10 per cent from imported wet salted hides from which such waste material has been removed before shipment. It is also significant that the weight of pig iron as a proportion of the blast furnace charge, while one-fifth for home ores, is one-third for imported ores, for only the richer ores are brought in from abroad.

These examples have been selected in order to demonstrate the value of Weber's proposition that weight-losing industries are drawn to their raw materials and they also serve to demonstrate that the strength of the tie varies with the amount of the loss. It is one thing, however, to observe its operation in certain selected cases and an entirely different thing to dignify it as a universal law. I propose now to turn to the full range of British industry and to examine Weber's hypothesis in this wider context. I shall draw my data from the 1948 Census of Production, the most recent full census, later censuses being partial and of sample particulars only. Values of materials and of production have been recorded in every census, but the first census which required weights was that of 1935. The 1948 Census records these weights more fully than the 1935 Census. But, while virtually complete for many industries, the particulars of weight are incomplete for many others. Even for industries as well documented statistically as iron and steel some weights are unrecorded. For other industries no weights whatever are given for production and weights are given for only a fraction of the materials. In tailoring and dressmaking, for example, production is recorded only in terms of values and of number of articles, and materials are recorded only in terms of square yards. No estimation is possible for production but some estimation is possible for materials: nevertheless, it can be no more than rough approximation, for cloth varies considerably in weight per square yard. The employment of the returns of the Census of Production for the examination of Weber's thesis, therefore, is full of difficulty. Should the weight of the unrecorded items be estimated on the assumption that the ratio of weight to value is the same for the unrecorded as for the recorded items? I have attempted estimation along these lines for every industry for which it is at all possible, but I am convinced that the estimations are of little value for the purpose of testing Weber's hypothesis. Rough approximations are valueless: order of magnitude is not enough, for the problem involves not single quantities but the relation of two quantities to each other. Moreover, the premise that the ratio of weight to value is the same for the unrecorded as for the recorded items is inaccurate, for it is usually the items whose weights are substantial for which weights are actually recorded and the items for which weights are unrecorded are usually those relatively higher in value in proportion to weight. This is not always true, but whether such methods overestimate or underestimate weight it is clear that at best they can only be approximate and that at worst they may distort the ratio completely. I have been compelled, therefore, to ignore those industries for which a substantial proportion of their materials or of their production is recorded only in terms of values or of quantities of such a kind that calculation of weights is impossible. Some others cannot be included as they are merged in the census classification with other trades of different character: thus, butter and cheese making is combined with milk bottling. Sixty-five industries remain. They range widely from those whose product is less than 20 per cent of weight of materials (including coal) to those whose product is more than twice the weight of materials (including coal).

They have been tabulated in Table 20.3 and classified by observation of each industry in turn whether in fact they display location at materials, such as blast furnaces, whether in fact they display partial location at materials and partial location independent of materials, such as leather tanning, or whether in fact they display location entirely independent of localised materials, such as bread baking or beer brewing. It will be recalled that Weber regards water as ubiquitous and, therefore, not a localised material. The tabulation shows that all industries located at materials have a value below 100, that is a material index of more than one. Those industries with a value over 100 and, therefore, with a material index of less than one are located either wholly or partly independently of materials. Weber's watershed has some validity. But industries located at materials are dispersed along the

TABLE 20.3

LOSS OF WEIGHT DURING MANUFACTURE

Weight of product as per cent of weight of material (including coal)

Industries	*10–19*	*20–29*	*30–39*	*40–49*	*50–59*	*60–69*	*70–79*	*80–89*	*90–99*	*100–109*	*110–199*	*200+*	*Totals*
Located at materials	2	2	1	1	2	3	3	5	3				22
Located partly at materials		2		2	2	1	1	1		3			12
Not located at materials		3	3	2	1	6	2	2	1	3	3	5	31
Totals	2	7	4	5	5	10	6	8	4	6	3	5	65

entire range of values below 100 and industries located independently of materials are likewise dispersed along the entire range of values below 100 with the single exception of that below 20. It may be inferred that the material index provides us with a tool of analysis but that it is a blunt tool and is effective only at the very extremities of the classification. For industries within the very wide range from a material index of one to a material index of five this index offers no certain guide to whether an industry is or is not located at its materials. This middle range comprises no less than 49 out of 65 industries and of these 20 are located at materials, 20 independently of materials and 9 partly at materials. It classifies together, for example, blast furnaces and cotton spinning, the one located firmly at the source of ore and coal and the other located with complete independence of the source of raw cotton. This is inclusive of coal in all cases.

Weber's inclusion of coal with industrial raw materials, however, confuses the issue for all industries except those, like blast furnaces and briquette plants, for which coal is a raw material in the real sense and not only a source of power. A high coal index, in fact, indicates tie to coalfield and not tie to raw materials as such. For the geographer the distinction is of great importance. The textile industries, for example, use considerably greater quantities of coal than of textile raw materials; they may be located on coalfields, but they are certainly not tied to their sources of wool, or cotton, or raw silk. In view of this I have recalculated loss of weight for each of the same sixty-five industries by excluding coal from materials (Table 20.4). The material index of one does not appear to have the same validity as when coal is included, but the two industries located at materials and yet with an index below one are both capable of rational explanation. One is the preserving and bottling of fruit and vegetables which, even if there be a gain in weight, must be located where such fruit and vegetables can be handled in fresh condition, and the other is oils and greases which should perhaps have been omitted from consideration as the weight of its raw materials is under-recorded. This second tabulation narrows the middle ranges and increases the range of effectiveness of the tool at the extremities. Of the industries with a material index of over 1·4 (that is, with weight of product less than 70 per cent of weight of localised material) most are located at materials, 6 wholly so, and only 2 are independent of materials.

TABLE 20.4

LOSS OF WEIGHT DURING MANUFACTURE

Weight of product as per cent of weight of material (excluding coal)

Industries	*Under 30*	*30–59*	*60–69*	*70–79*	*80–89*	*90–99*	*100–109*	*110–119*	*120–199*	*200+*	*Totals*
Located at materials	2	1	3	4	5	5	2				22
Located partly at materials		1	1	1	1	3	1	3	1		12
Not located at materials		1	1	4	6	3	2	3	4	7	31
Totals	2	3	5	9	12	11	5	6	5	7	65

Of the industries with a material index of under 1·0 (that is, with weight of product more than weight of localised material) most are located independently of materials, 16 wholly so, and only 2 are located at materials. But the middle ranges (even when defined thus less rigidly than in the interpretation of Table 20·3) still include half of the industries in the table, and of these 14 are located at materials, 13 independently of materials and 5 partly at materials. In these middle ranges the material index jostles with other conditions to determine location.

Is it possible to devise other tests of the location of industry on the site of its materials in addition to loss of weight during the course of manufacture and, in particular, is it possible to sort out industries in the middle ranges of the loss-of-weight classification for which that classification gives us so little help? Industries may have the same percentage loss of weight during manufacture, but may employ very different aggregate weights of materials. The test I wish to consider is the weight of materials per operative,[2] and the data I shall use are those of the 1948 Census of Production. The weight of materials processed by each industry varies, of course, with the size of the industry and it is essential, if one industry is to be compared with another on a common list, to relate them to some common denominator. The number of operatives has been used for this purpose. Numbers of operatives returned are not, of course, strictly comparable, for there is the difference between adult and juvenile, male and female, good worker and poor worker, mechanised industry and handicraft industry. But they are at least fully recorded.

The industries in the middle ranges of the classification according to loss of weight have been tabulated according to weight of material per operative (Table 20.5). They display wide range in weight per operative per annum from 1450 tons in class A to 3 tons in class M.[3] These were then classified as previously according to whether by observation they display location at materials, location partly at materials or location independent of materials. The result is striking, for these industries in the middle ranges of rank are arranged into two groups with little overlap. Those with high weights per operative are located at materials, those with low weights per operative are not located at materials. The break between them comes at 30–40 tons of material per operative per annum. This is exclusive of coal.

TABLE 20.5

WEIGHT OF MATERIAL PER OPERATIVE

Classification according to weight of material (excluding coal) per operative, greatest at A, least at P

Industries	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Totals*
Located at materials	4	3	2	3	1					1							14
Located partly at materials						1	2		1		1						5
Not located at materials					1	1	2		3	2		2	2				13
Totals	4	3	2	3	2	2	4	0	4	3	1	2	2	0	0	0	32

These industries are those in the middle ranges of the classification according to loss of weight (cf. Tables 20.3 and 20.4).

If weight of material per operative is so effective a tool in sorting out these industries of the middle ranges of the loss-of-weight classification with regard to the observed facts of location at materials, may it be equally effective over the whole range of industry? I have examined it for the 65 industries which have been considered previously for the operation of loss of weight (Table 20.6). I have examined it also for the 125 industries for which there are records adequate for the purpose (Table 20.7). It is an effective tool in both cases, though the amount of overlap is greater than for the smaller sample. The boundaries between the two groups, those located at materials and those not located at materials, are blurred, but the area of uncertainty is very much narrower and more sparsely populated than in the classification according to loss of weight. The area of overlap is, of course, greatest for those industries located partly at materials and partly independent of materials, like the tanning of sole leathers, at

TABLE 20.6

WEIGHT OF MATERIALS PER OPERATIVE

Classification according to weight of materials (excluding coal) per operative, greatest at A, least at P

Industries	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Totals*
Located at materials	6	3	5	4	1		1			2							22
Located partly at materials		1	1		2	1	2	2	1		2						12
Not located at materials					2	3	3	3	5	4	2	3	4	1	1		31
Totals	6	4	6	4	5	4	6	5	6	6	4	3	4	1	1	0	65

These industries are those which appear in Tables 20.3 and 20.4.

TABLE 20.7

WEIGHT OF MATERIALS PER OPERATIVE

Classification according to weight of material (excluding coal) per operative, greatest at A, least at P

Industries	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Totals*
Located at materials	7	6	6	7	2		1			2							31
Located partly at materials	1	2	1	1	2	2	2	3	1		2						17
Not located at materials			1		4	4	5	5	7	6	6	8	8	8	8	7	77
Totals	8	8	8	8	8	6	8	8	8	8	8	8	8	8	8	7	125

All industries for which information is available.

materials, and of upper leathers, independent of materials, or like earthenware and refractories, at materials, and china, independent of materials.

It would thus appear that weight of material per operative is a sharper tool than loss of weight during manufacture for sorting out industries into that group located by materials and into that group located independently of materials. It is a sharper instrument partly for mechanical reasons, the nature of the returns of the Census of Production and the greater spread of values. But there is a more fundamental reason. Loss of weight has significant locational effects only when it is combined with large weight per operative, for variations in transport costs are substantial enough to affect location only if weights handled are large. A combination of large loss of weight and large amount of material per operative ties an industry in location to its raw materials; a combination of large gain in weight and small amount of material per operative frees an industry from location at its raw materials; where the two conflict, as in many engineering trades which produce large quantities of metal scrap and which therefore display substantial loss of weight during manufacture but which handle only small weight of material per operative, it is the weight of material per operative which is the more significant in its effects on locational ties to material. Engineering industries in the mass are now by no means tied to material.

The analysis has been of contemporary conditions using contemporary data. There is reason to believe that particular industries have changed in the strength of their locational ties to materials with the course of time. There is a large number of industries not now tied to materials but which seem originally to have been tied to materials often on the same sites which they now occupy. Examples are the older finished iron trades, such as chains and hardware, tools and cutlery; the mechanical engineering trades, such as textile machinery and locomotive building; and the textile trades, although the particular textile material may have changed. These all employ only relatively small weights of material per operative per annum, much smaller than is commonly realised, fifteen tons for chains, nails and screws, less than one ton for cutlery,

five tons for textile machinery, six tons for locomotive shops, three tons for cotton, two and a half tons each for wool and for linen, less than one ton for hosiery. Each of these has a well-defined regional distribution but their materials are no longer drawn wholly locally, if indeed they are drawn locally at all. Some would doubtless argue that they present examples of industrial inertia. The phrase is not an apt one if by inertia is implied maladjustment to present conditions. There is no alternative site within Britain where they would be better placed. Even small weights of material per operative were significant when transport was rudimentary and expensive, but they are not significant today. The small weights of materials which they use can now be drawn in readily from elsewhere by the present transport system. It is, indeed, open to question whether some of these industries, particularly the textile, were ever as firmly located by raw materials as is commonly supposed. They have now a very solid balance of advantages in their present location, established labour skills, external economies as a result of the aggregation of like with like, and, although both materials and markets are now distant, there the industries remain and they move only if the solid advantages which have kept them so long begin to evaporate. The shortage of labour in the West Riding has led some wool textile firms to migrate to districts where labour is available, even if it is raw and untrained. While some industries are less tied to raw material than formerly, there are some others, fewer in number, whose tie to raw materials is now stronger than previously. Grain milling, newsprint, tailoring and dressmaking are examples; they have become concentrated in the ports and in the textile districts respectively. In these instances the condition responsible for the change is the greater weight of material per operative as a result of more intense mechanisation, thus strengthening the tie to materials.

There is another set of industries whose location pattern with regard to materials is subject to change. Large loss of weight ties industry to raw material provided it be linked with large weight of material per operative, and this large loss of weight is due to the production of large quantities of waste material for which no economic use can be found. When an economic use opens up for these waste products the most economic site may be changed thereby. In some cases the economic use is at too low a value to carry any locational significance. Thus sugar-beet pulp, even

Fig. 20.4 Coke ovens, 1952

when prepared for sale, is only a third of the value of the unrefined sugar; and basic slag, even when ground, is only a fourteenth of the value of pig iron. But in a few cases it is sufficiently high to be significant. When coke was made wholly in beehive ovens and the coke-oven gas ran to waste, then the pit head was the most suitable site for coking plant for there was a substantial loss of weight in a bulky material. But when coke-oven gas proved to be usable for heating in steel plant, then there was advantage to be gained by locating coking plant adjacent to steel works. As coke is used in

greater quantities in blast furnaces than in steel works, it was at integrated plants with both blast furnaces and steel mills that such coking plant was set up. Coke plants are now, therefore, located not only at pit head but also at integrated blast furnace-steel plants; some of the former pass their waste gas into town mains (Fig. 20.4).

I come now to coal as a location factor. I have already pointed out that coal and materials should be distinguished. We are all familiar with the powerful locational influence of coal during the first half of the nineteenth century when steam-power, generated by coal-fired boilers, created a seething mass of industrial plants on the coalfields. Coalfield and factory industry were coincident. The metalworking, mechanical engineering and textile industries, the last using larger quantities of coal than of materials, congregated on the coalfields and there were relatively few plants of these kinds elsewhere. As boiler efficiency was low, the amount of coal consumed per operative was higher than it is today, and transport of coal away from the coalfield was then both expensive and trouble-some. The coalfield monopoly of factory industry has subsequently been infringed. This has been due not so much to the emigration of these industries from the coalfields, though some blast furnaces and steel plants have thus moved, as to the creation of new industries on new types of site as the British economy has become diversified. The railway made it possible to move coal more readily about the country and it became possible for industrial plants to be set up away from the coalfields, whether they were run by steam-power or by electricity. At first, electricity was generated expensively by coal in small plants in consuming districts, for transmission of electric current was only over short distances. Transfer to electric power was thus retarded. The completion of the Grid, permitting current to be transmitted across country, has accelerated such transfer, and acceleration is likely to be increased still further by the concentration of output which is now in progress on low-cost sites in gigantic plants located on the Trent near to the largest and most cheaply worked coalfield in the country.

Weight of coal per operative has been plotted against weight of material per operative in terms of rank in each case and expressed

TABLE 20.8

WEIGHT OF MATERIAL PER OPERATIVE AND WEIGHT OF COAL PER OPERATIVE

Classification according to weight of material per operative	*Classification according to weight of coal and coke per operative, greatest at A, least at P*																
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Totals*
A	3	2		2								1					8
B	1	1	3	1	2												8
C	3	3		1	1												8
D		1	1			1	1					1	1		2		8
E			1		1	1	1		2		1	1					8
F	1		1	2	1			1		1					1		8
G			2			1	1	1	2	1							8
H		1		1	1		2			1			1	1			8
I							1	1		3		2				1	8
J				1	1		1	1	1	1	2						8
K						1		1	1	1	1			1	1	1	8
L					1	1		1		1	1	1	1			1	8
M						3	1	1			1		2				8
N												1	1	4	2		8
O								1			1	1	1	2	1	1	8
P									1			1		1	2	3	8
Totals	8	8	8	8	8	8	8	8	7	9	7	9	7	9	9	7	128

in tabular form in Table 20.8. In many industries there is a reasonable degree of agreement in rank. There is absolute agreement in 17 per cent and little divergence in another 29 per cent of industries. The metal and engineering trades are examples of close agreement in rank, whether they be of high rank, as blast furnaces, of middle rank, as foundries or mechanical engineering, or of low rank, as electrical engineering. But there is substantial divergence in half the examples and the two classifications are often widely at variance. For 29 per cent they are a full quartile removed. Some with a high rank in weight of material per operative use little power of any kind, such as fish curing or textile converting, and some, if they use power in quantity, employ mainly electricity, such as saw-milling. They are located by materials but are not at all attracted to the coalfields. Others with a low rank in weight of material per operative have a conspicuously higher rank in weight of coal per operative. The textile industries are almost wholly of this kind. They use small quantities of material per operative, as I have shown, but they subject that material to a succession of mechanical processes which require substantial quantities of power, and that power is still steam-power generated from coal on the mill site. The tie of these textile industries to a coalfield, however, is relative and not absolute, for the coal they use still places them only in the middle rank and their coalfield site is largely a relic of the early nineteenth century. For the tie to a coalfield to be absolute the quantities of coal must be large and they must exceed the quantities of the raw materials, unless those raw materials are themselves located in the coalfield. Few industries fulfil these conditions today. It is primarily those industries processing coal in large quantities as a raw material which are tied to a coalfield. I do not wish to be misunderstood. Many industries derive positive advantage from location on a coalfield, but these advantages, it is necessary to repeat, are relative and not absolute.

Fig. 20.5 Open-hearth steel furnaces, 1953
Note: A 'heat' is the full cycle of operations for the production of steel. The number of such cycles per working week varies from place to place, with the type of furnace and with the materials used.

In conclusion and by way of summary I will endeavour to demonstrate the validity of the conclusions of this paper by reference to the location pattern of a particular sequence of industries manipulating a particular material in a more and more elaborate form. The sequence is, first, blast furnaces; second, steel mills; third, tube mills; fourth, chain, nail and screw mills; fifth, textile machine shops; sixth, motor vehicle manufacturing shops. The weight of the finished product as a percentage of the weight of materials is for blast furnaces 27 per cent, for steel mills 81 per cent, for tube mills 84 per cent, for chain, nail and screw mills

74 per cent, for textile machine shops probably no more than 50 per cent, but for motor vehicle shops it is not possible to make a calculation owing to inadequacy of evidence. For the first three items in the sequence there is less and less loss in weight as the material becomes more highly manufactured, but then loss of weight increases again owing to the production of scrap in machine shops. The weight of material per operative is for blast furnaces 1447 tons, for steel mills 117 tons, for tube mills 38 tons, for chain, nail and screw mills 15 tons, for textile machine shops 5 tons, and for motor vehicle manufacturing shops 7 tons. There is a consistent decrease in weight of material per operative as the material becomes more highly manufactured. The increase in motor vehicle shops over textile machine shops is due to methods of line assembly and to prefabrication elsewhere, both of which have increased weight of material per operative in the assembly shops themselves. Weight of coal per operative varies in the same sequence and more regularly, for coal per operative is less in the motor vehicle shops than in the textile machine shops. As these quantities change, so also does the nature of the location pattern. Blast furnaces are tied to sources of ore or coal, being mainly at the source of ore; steel mills are partly at the site of blast furnaces and partly elsewhere, being mainly on the coalfield (Figs 20.5 and 20.6); textile machine shops are in textile districts; and motor vehicle manufacturing shops are in the English Plain removed from blast furnaces and steel works altogether, with the single exception of the Ford works at Dagenham.[4] The distribution of steel works overlaps the distribution of blast furnaces; the distribution of rolling mills overlaps the distribution of iron works and steel works; the distribution of engineering shops overlaps the distribution of rolling mills. But no two of these distributions coincide. Each stage overlaps the preceding, but the terminal points, smelting and engineering, are only rarely associated in the same firm on the same site. The more and more elaborately manufactured a material becomes, the more completely do the mills or shops which manipulate it become divorced in location from the plants which handle the original raw material. A sequence of this kind displays the full range of locational shift from a location tied to raw material to a location wholly independent of raw material.

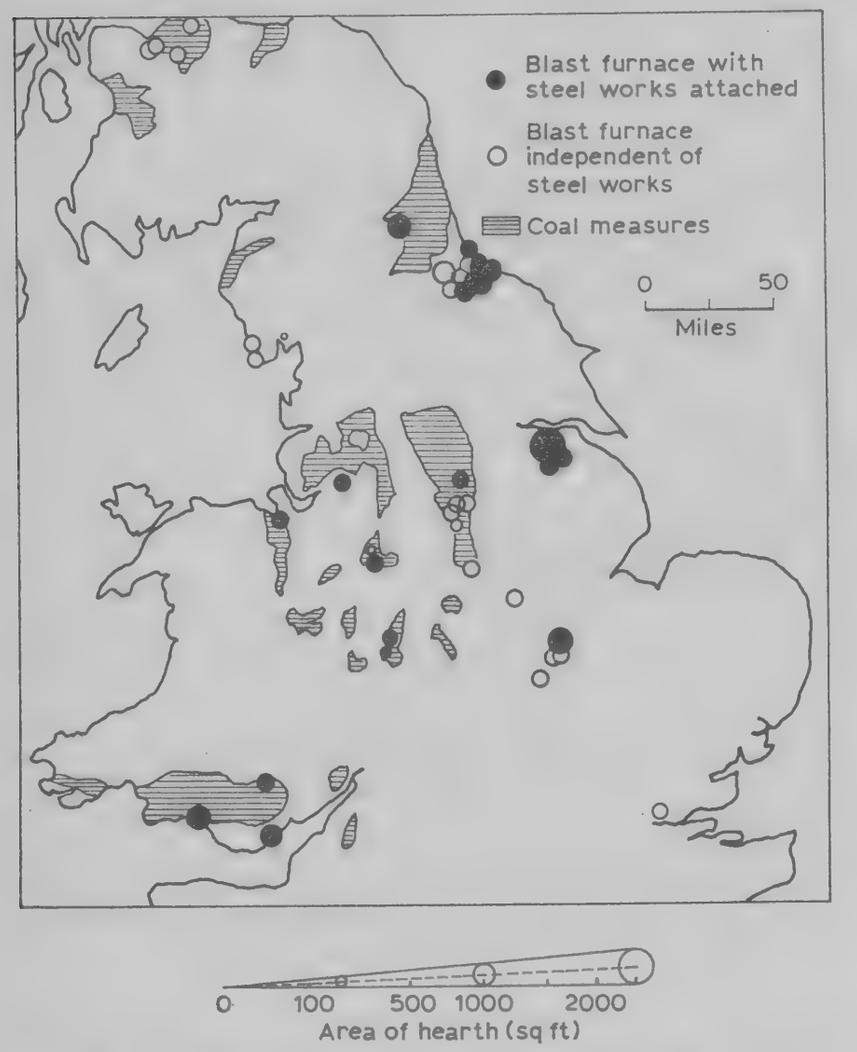


Fig. 20.6 Blast furnaces and open-hearth steel works, 1953

Notes

1. *Transactions of the Institute of British Geographers*, vol. 21, pp. 1–18.

2. I have employed this before in 'Mobility in the Location of Industry in Great Britain', *Advancement of Science*, vol. 6 (1949–50) 115–17 and in Appendix C of *An Economic Geography of Great Britain*, 2nd ed. (1953). It has been pointed out to me by Mr Brian Law that Weber's labour coefficient is not unlike my weight of materials per operative, except that it refers to costs of labour, not to quantity of labour irre-

spective of cost. This paper does not include an examination of costs of materials, of value of product or of labour costs of conversion. I have considered costs of materials and value of product in the two papers listed above, but have not yet examined labour costs of conversion. This paper confines its attention to weight: the analysis of costs must be left for a further occasion.

3. The classification into sixteen classes (A–P) is the same in Table 20.5 as in Tables 20.6 and 20.7. The three tables can thus be directly compared. There are eight industries in each of the sixteen classes. Table 20.7 embodies the full list of 128 industries, less three whose particulars are too incomplete for inclusion. Weight of material per operative is greatest in class A and least in class P.

4. The middle members of the series, tube mills and chain, nail and screw mills, display interesting examples of the problems of location presented by industries in the middle ranges of classifications both on the evidence of loss of weight during manufacture and on the evidence of weight of material per unit of labour. Both kinds of mill were originally located in old iron-working centres with ore and coal both present. Most plants still remain here despite changes in the provenance of raw materials and despite changes in loss and cost of transport; nevertheless, the largest tube mill is now on an ore-field as part of an integrated blast furnace and steel plant.

Smith found then that, while the material index discriminated reasonably well industries located at the raw material end, such as sugar beet processing, from those otherwise located, the correlation was far from perfect. His results were clearer when the weight of coal was excluded, but dissatisfied with results obtained using the material index, he went on to look at other indices that might be helpful in separating out industrial locations. As he demonstrated, these indices were at least as discriminating as Weber's index, and where values for Smith's indices were quite high, certainly strong material orientation in location was evident.

WEBER'S MODEL AND REALITY

If Smith's work hinted at weaknesses in Weber's concept of the material index as an indicator of industrial orientation and criticised Weber for including the use of coal in the weight index because it confuses raw material with fuel orientation, it must of course be apparent that the basic least-cost location or minimum energy model is far from being a total explanation of industrial location in a real-world situation. To begin with, it is as unreasonable to assume perfect competition and given markets in this model as it was in von Thünen's. Besides which, many factors that influence industrial location in reality are ignored in this model as they were in the 'isolated state'. It will be recalled that in the last chapter Wolpert stressed the importance of the farmer's imperfect knowledge of his total situation as an agricultural producer and his propensity to behave in a non-rational way. Equally, although Weber ignored the possibility, the industrialist will certainly lack perfect knowledge of his situation. Weber assumed optimal behaviour, but as J. G. March and M. A. Simon have pointed out with reference to human behaviour in business organisations:[1]

Most human decision-making, whether individual or organisational, is concerned with the discovery and selection of satisfactory alternatives; only in exceptional cases is it concerned with the discovery and selection of optimal alternatives. To optimise requires processes several orders of magnitude more complex than those required to satisfice. An example is the difference between searching a haystack to find the *sharpest* needle in it and searching the haystack to find a needle sharp enough to sew with.

Note

1. *Organisations* (New York, 1958) pp. 140–1.

The idiosyncratic nature of human judgement, which is also presupposed by an approach which is not entirely rational, plays a part in industrial location. As Greenhut and Colberg have argued in 'Factors in the Location of Florida Industries' (1962), given the same plant to locate, no two industrialists would judge the possibilities for profit maximisation in the same way.

One enterprise and enterpriser may be conservatively formed; another may be more daring in plans and in deeds. Opposite choices may be made by comparable firms because the material make-up of their personnel differs. We might find that one business unit wants sure bets, while another seeks the lure of greater gains generally commensurate with greater risk. Or, the financial position of one may be such that only a small risk can be taken, whereas another can afford a large risk. In effect, we reach the result that the decision-making process may yield opposite choices, even though the parties of interest claim their desire for the same end: namely, greatest pecuniary profits.

Personal considerations in the location process, whether partly economic or purely social, will also play their part. There is evidence that the final decision to set up a business in a particular place may be influenced by climate or recreational facilities such as golf or sailing. In her study of the 'Post-War Industrial Development of Plymouth', J. C. Braithwaite[1] *has pointed out that the fact that the city is situated in a pleasant part of Britain has been an important factor in enabling companies to recruit highly qualified managerial staff.*

As individual preferences can be a location factor, corporate decisions by governments can also play an even more important role in the spatial placing of industrial concerns. Although the Weberian model takes no account of such interference at state level, governments can take actions in this field on the grounds of social benefit and national interest. Socialist or Communist governments with their emphasis on planning and control are more likely to exercise their influence here. F. E. I. Hamilton has examined this situation in his paper 'Location Factors in the Yugoslav Iron and Steel Industry'.[2] *In this he maintained that political factors had been of paramount locational importance. Between the wars, the industry suffered from the economic and political problems of integrating a weak and newly independent state. But since 1945 rapid expansion at specific sites had resulted from the application of socialist policies, though the influence of economic considerations were not to be entirely discounted. Since the end of the last war Hamilton noted three major trends:*

First, there has been an expansion of Slovenian plants, though at a rate below that of the whole industry. Second, the concentration of much new coke-battery, blast furnace and steel capacity at Zenica in Bosnia has made this the major metallurgical centre. Third, three plants have been built in new locations, namely, at Sisak (Croatia), Ilijaš (Bosnia), Nikšić (Montenegro); a fourth is under construction at Skoplje (Macedonia) (see Fig. 20.7).

These developments result from the application of principles of Communist policy for locating plants near their raw material supplies, or in overpopulated, economically backward areas.[3] Yet the locations chosen also reflect the impact of changed Yugoslav relations with the world outside and the adaptation of the industry under these conditions to specific characteristics of the country's political and economic geography.

Expansion of Slovenian steelworks followed from economic factors emphasised in the Five-Year Plan, 1947–51. It aimed at 'reconstructing existing iron and steel plants' to integrate their production. Far removed from material sources, these plants demonstrated the effects of inertia. Their expansion has resulted from the following factors: the urgent need for metal supplies in industrialising the country and for building metallurgical indus-

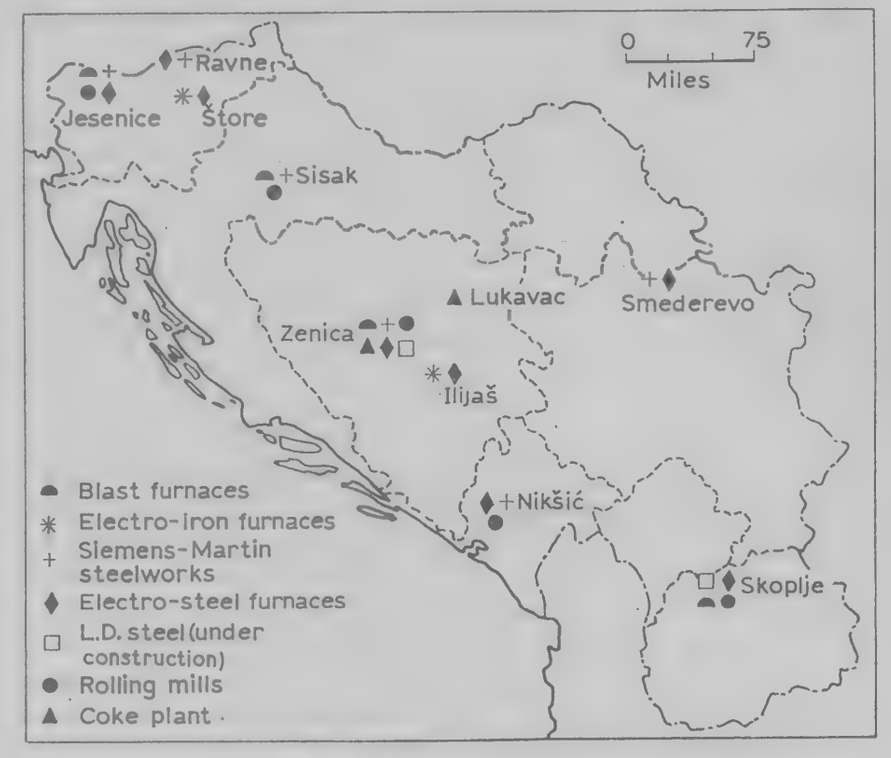


Fig. 20.7 Location of plants constructed between 1945 and 1963

tries in other regions; the local availability of skilled labour; the possibility of manufacturing high-quality (including electro) steel using local scrap and water power; the inability of Yugoslavia to invest in the relocation of the plants; and their proximity to markets in the north-west for steel plates, sheets, wire and quality steels. Economic factors thus overruled any strategic considerations which might have required the removal of steelworks from the north-western frontier with 'capitalist' countries, after 1945. The importance of this area has declined, however. Its peripheral posi-

tion leaves it strategically vulnerable,[4] isolated from materials and increasingly distant from developing markets in eastern and southern Yugoslavia. Production costs are high, and the advantages of other areas greater. For example, in 1960 the Jesenice plant obtained most materials from Bosnia – 300,000 tons of ore from Ljubija (180 miles by rail), 160,000 tons of coke from Lukavac

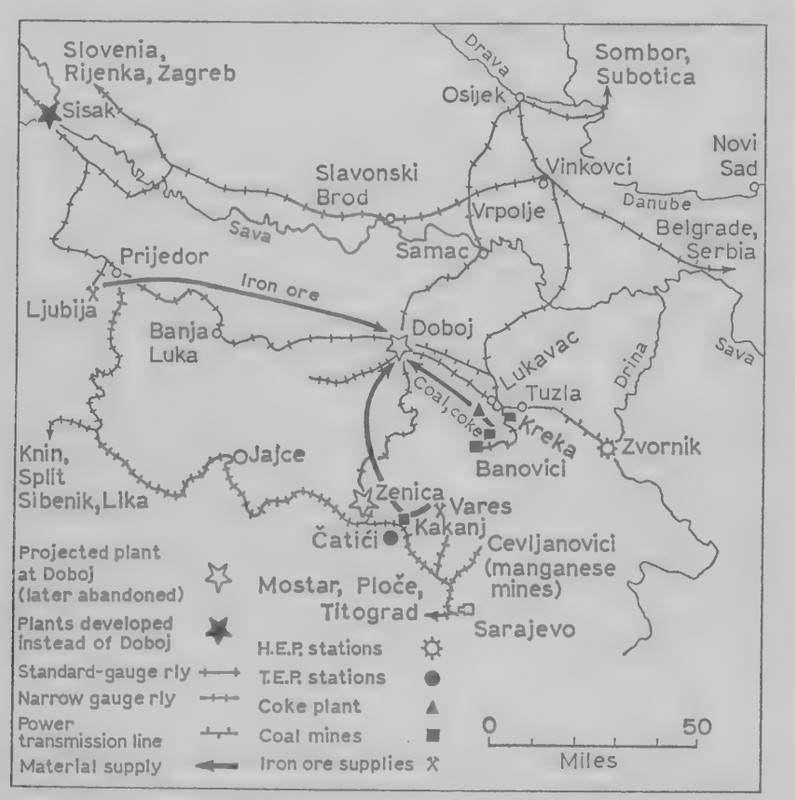


Fig. 20.8 The Doboj project: the plan for industry in 1947

there. Doboj offered a good location for the plant: an excellent assembly-point for iron ore from both Ljubija and Vareš, hauled along newly constructed standard-gauge railways,[5] for coke from Lukavac, and for local limestone (Fig. 20.8). Doboj was also well situated for dispatching steel to major engineering markets in the

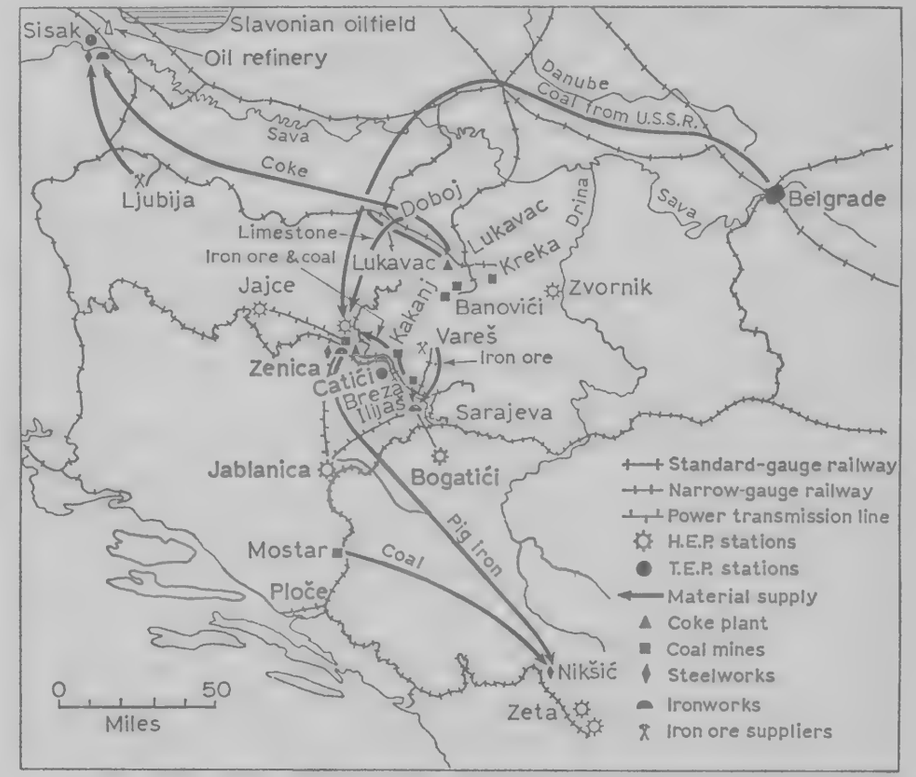


Fig. 20.9 New plants constructed after the Cominform break, and the changed pattern of material movement

(310 miles) and 75,000 tons of coal from Banovići (320 miles). Only scrap and coal from Trbovlje were supplied locally, so that steel production costs were $201·6 per metric ton.

The Five-Year Plan also indicated a major change in the industry's distribution. It required 'the construction of one new metallurgical combine in the Bosna Valley', to be located at Doboj, not at Zenica. To develop most new capacity in central Bosnia was only logical in view of the minerals and power localised north-west and north-east, and to economic 'growth areas' in central, eastern and southern Yugoslavia.[6]

The Cominform Resolution had a profound effect upon these plans (compare Figs 20.8 and 20.9). By 1949, all supplies of credit, materials and equipment from Soviet Bloc countries had ceased; and there was the threat of invasion. The Doboj project was abandoned. Another plan was drawn up for developing 'key projects', defining their character and location.

The new plan, which required the building of plants at Zenica, Sisak, Nikšić and Ilijaš, was a response to four factors: first, Yugoslavia's enforced dependence upon her own financial resources; second, a similar dependence upon her own natural resources; third, new strategic needs; and fourth, problems of the country's political integration under new conditions.

First, the acute shortage of capital demanded locations where the investment costs of plant and ancillary services were lowest, i.e. at Jesenice and Sisak in particular. Capital was a major factor in the abandonment of the undeveloped site at Doboj in favour of Zenica where plant, transport, power, water and housing facilities already existed.

Second, by throwing Yugoslavia into much greater dependence upon her own natural resources, the Cominform break speeded the introduction of electric furnaces. This stimulated considerable expansion at existing plants (in Slovenia) and the development of new plants in areas (Bosnia and Montenegro) of expanding hydro-electricity output.

Third, the threat of attack from the north and east resulted in the location of some capacity in strategically more isolated and more secure areas (in 1949) in mountainous central and south-western regions (Bosnia and Montenegro). The result was greater dispersion than had been planned previously. Hence the construction of new plants (not planned in 1947) at Sisak, Ilijaš and Nikšić. The importance of Smederevo was most affected by the Resolution. Before 1941, direct water[7] and rail links with Soviet Bloc countries for supplies of materials and distributing products encouraged expansion at Smederevo; after 1948, development was discouraged because of its proximity to the vulnerable north-eastern frontier.

Fourth, the regional distribution of new plants reflects the influence of changed internal political conditions. Between 1919 and 1944, the historical, political and cultural differences between the various Yugoslav peoples had led to increasingly bitter conflicts. These culminated in the Civil War of 1942–4. While Yugoslavia was securely integrated with the Communist Bloc, these national differences were subordinated to economic decisions of development. But post-war Yugoslavia had been reconstituted as a federal state in which all peoples had equal rights. The government believed there could be no such political and social equality without considerable 'national', i.e. regional, *economic* equality. It became urgent to apply this policy after 1948 when Yugoslavia became politically isolated and was subjected to intense propaganda from the Soviet Bloc in an attempt to revive nationalist feelings against the federal government. The latter countered the external political pressure working on latent internal jealousies by locating steel plants in national republics (Montenegro and Croatia) which lacked the heavy industrial basis for their economic development.[8]

Notes

1. *Transactions of the Institute of British Geographers*, vol. 45 (1968) 44.
2. *Economic Geography*, vol. 40 (1964) 57–61.
3. The Yugoslavs have never considered the market as a locating factor, owing to its overall weakness and its division into small north-western and north-eastern areas.
4. This was the view taken before modern weapons were developed.
5. Before the war no rail link existed between Ljubija and Doboj, and only a poor narrow-gauge line served Vareš and Zenica from Doboj. In 1949–50 new lines were built linking Sarajevo with Brod and Banja Luka through Doboj.
6. These are areas in which development in communications, mining, building and manufacturing was to proceed at rates much higher than elsewhere in the country.
7. Via the Danube upstream to Hungary and Czechoslovakia, and downstream to Romania, Bulgaria and the U.S.S.R.
8. Already by 1949 Slovenia, Serbia and Bosnia possessed such industries. A metallurgical plant was promised for later construction in Macedonia for these same political reasons; this is now under construction at Skoplje.

Outside countries which rely on Communist planning, freely elected governments can play their part in industrial location. In the United Kingdom this is achieved through the use of Industrial Development Certificates. Firms may or not be granted these depending on where they wish to set up new industrial plant. Development areas[1] will certainly be favoured, and in these further financial incentives may act as a magnet to firms.

In a recent paper on 'Industrial Development in Ireland' (1967), J. R. Tarrant[2] has discussed the role of government in that country as an influence on location.

For the purposes of industrial development the country is divided into three regions. The western and some of the central counties

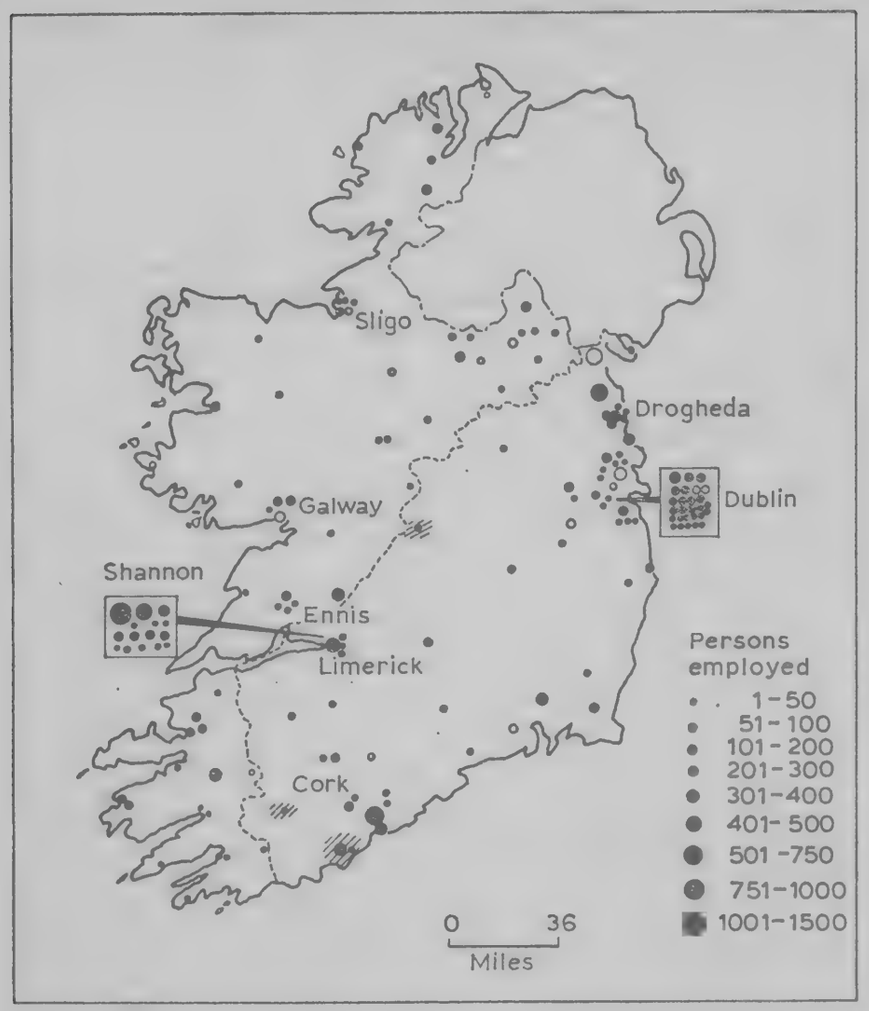


Fig. 20.10 The distribution of industrial employment in the Republic of Ireland established with foreign participation between January 1959 and December 1966.

The closed circles denote employment in new industries and the open circles employment provided by expansion between 1959 and 1966 of firms established prior to 1959. Areas to the west of the dotted line represent the development areas. Small enclaves, considered as part of the development areas, east of this line are indicated by the shaded areas.

(Fig. 20.10) are designated as a development area within which the Irish government is particularly anxious to see industrial expansion in order to limit emigration. Industries established within these development areas are eligible for a higher rate of aid than in the rest of the country. There is also the special case of the Shannon Free Airport Scheme where the advantages to industrialists are different from those in the remainder of the country.

The aid and encouragement given to industrialists falls under the four headings of capital grants, loans, grants towards the costs of training workers and tax reliefs. Although there are grants paid towards the costs of expanding and modernising existing plants, most of the available aid goes to the establishment of new industry, and firms with foreign participation have benefited considerably from this assistance. Up to two-thirds of the total costs of building and equipment in the development areas and one-half in the rest of the country are payable in the form of capital grants. Loans, both of long and short duration, are available to new industries through the State-sponsored Industrial Credit Company. These are granted at very favourable interest rates primarily for the acquisition of fixed assets. Grants are also obtainable towards the costs of training workers up to the total amount of their wages during the training period.

Tax reliefs include an initial allowance of 20 per cent of the cost of an industrial building and 50 per cent of the capital costs of new plant and machinery, excluding road vehicles. There are also annual allowances towards the same costs. As the Irish government is particularly concerned with the encouragement of exports, there are further advantages enjoyed by a new company exporting for the first time. Concessions amounting to 100 per cent tax relief on profits attributable to exports are allowed for ten years from the establishment of the export trade. This scheme will end in the year 1979–80. Companies starting in the export trade before 1970–1 have an additional period of partial relief, the length of which depends on the date of establishment.

An important part of the encouragement to foreign industrialists is the development of the Shannon Free Airport Scheme within which firms operating for export markets can import and export without customs restrictions and with certain other benefits. Grants of up to 50 per cent of the costs of machinery and 100 per cent of the costs of building are payable as well as the total costs incurred

through worker-training programmes. Profits arising from exports are exempt from duty until 1983. In addition, all imports of machinery, raw materials or other equipment are duty-free provided they are re-exported or retained within the limits of the customs-free zone. In addition to these financial and fiscal incentives there are other encouragements to industrialists. Standard factory premises can be rented or purchased at Shannon and, if these are not satisfactory, in certain circumstances a factory can be supplied to meet the industrialist's own requirements. Similar factories will shortly be available at Waterford and Galway. There are houses and apartments at Shannon for renting or purchase or alternatively sites can be reserved for personnel who wish to build their own houses.

The Shannon development was originally started to revive employment which had begun to fall off in the mid-1950s when jet aircraft were introduced on the Atlantic route and Shannon Airport was no longer needed as an important stopping point. A new jet runway became operational in 1960 and air transport facilities are available for the industries located on the Shannon estate. Those industries making products with a high value-to-weight ratio, readily adaptable to air transport, are therefore especially attracted to the Shannon location.

The results of these industrial incentives can best be measured by an examination of the distribution of employment offered by newly established industries. Industry established between January 1959 and December 1966 with foreign participation accounted for about 73 per cent of all new enterprises and 80 per cent of the estimated capital investment in new industries during that period (Fig. 20.10). There are two important concentrations of new employment, one in the east-coast zone stretching from Dublin north to the border, and the other in the area of Ennis, Shannon and Limerick. Most of this new industry is outside the development areas, although Shannon is a special case. There are less important centres in the Cork Harbour area and at Galway, Sligo and Tralee. Although there is a heavy concentration of new industries in the Dublin region, an examination of the employment offered shows that this area is not quite so dominant as it appears as most of the new industries have only a relatively small labour force. The three most important countries participating in Irish industry are the United Kingdom, Germany and the United States, while France, Holland, Denmark and South Africa are of lesser importance (Table 20.9). The British share has declined since 1958 when seven out of the total of ten new establishments with foreign participation were British.

The type of industry established over the last ten years reflects both the diverse needs of a country lacking any large-scale

TABLE 20.9

FOREIGN PARTICIPATION IN IRISH INDUSTRY

The number of new firms established with foreign participation between 1959 and 1966

Year of establishment	*Foreign participation*								
	U.K.	*Germany*	*U.S.*	*Holland*	*France*	*Denmark*	*S. Africa*	*Others*	*Total*
1959	4[a]	2	4	1	—	—	—	—	11
1960	5[b]	6	6	1	—	1	—	3	22
1961	7	6	4[c]	3	—	—	1	—	21
1962	6	6	—	1	2	—	2	1	18
1963	7	5	5	—	1[d]	2	1	—	21
1964	5	4	4	—	—	2	—	—	15
1965	4	5	3	—	3	—	—	2	17
1966	11	5	7	1	1	—	—	2	27
Total	49	39	33	7	7	5	4	8	152

[a] Includes 1 with U.S.
[b] Includes 1 with U.S. and Germany.
[c] Includes 1 with Holland.
[d] Includes 1 with Italy.

industrial development and also the 'footloose' character of the concerns attracted by government incentives to relatively isolated areas. For example, there is the only major oil refinery in the Republic at Whitegate, Cork Harbour. This was established in 1959 by the Irish Refining Company in association with the Caltex Oil Company, Esso Petroleum Company and Shell-Mex and B.P. Limited. There are also industries associated with the agricultural sector of the economy. The Wexford Creamery, established with help from 'Cow and Gate' of Britain, is one example of many milk-processing plants throughout the dairy-farming area of Ireland. There is also a large fertiliser factory at New Ross, established with Dutch participation in 1962. The 'footloose' industries cover a wide range of products, from small electrical components at Shannon to bathroom scales at Sligo and ball-bearings at Tralee to ladies' overalls and dresses at Letterkenny in County Donegal.

The existence of the special development areas seems to have made little difference to the overall distribution of industry. The area to the west of a line from Galway to Cork, including the parts of the development areas in counties Clare, Limerick and West Cork, has a relatively large number of new industries, but, on the other hand, the whole central and north-western part of the country has relatively few. Many of the small factories in the central counties of the country are milk-processing plants and apart from these there is little industrial development. Perhaps the most noticeable gap is in the Wexford–Waterford area in the south-east of the country. Waterford, with a population of 28,216, is the fourth largest town in Ireland. It is this lack of development in the south-east which has encouraged the Irish government to set up an industrial estate in Waterford, providing standard factory premises at a reasonable rent for new industrial enterprises.

Despite the extra incentives, the west coast, with the exception of the south-west, Galway and Sligo, has received little of the new industrial development since 1959. On the other hand, the planning of the Ennis, Limerick and Shannon area as an alternative magnet for industrial employment to the Dublin region seems to have met with considerable success. This has not been achieved without heavy government expenditure. In the financial year ending March 1966 the total government grants paid to the Shannon estate, plus the deficit of expenditure over income, amounted to £527,791. This figure includes the financing of tourist promotion for the benefit of the airport.

Notes

1. Areas with high unemployment which could benefit from industrial development.
2. *Geography*, vol. 52, pp. 404–7.

Besides those factors which Weber ignored in his model, other assumptions that are essentially 'unreal' were made by him. For example, like von Thünen, he also assumed that freight rates were proportional to distance. Alexander's paper quoted in the last chapter is sufficient proof of this fallacy and shows that they tend to be convex and stepped with distance. Often, weight for weight, raw material charges are lower than those for products.

It is also apparent in the Weber model that his industries A, C and E were purely theoretical since it would be difficult to conceive of industries that use only one material input. He assumed unlimited labour supply and fixed local differences in labour costs, while in reality labour can move between areas and industries, and in regions of rapid growth shortages of workers can occur.

Writers concerned with the Weberian model have added further criticisms. W. H. Dean[1] *has suggested that Weber overemphasised the importance of weight-losing materials on location when in fact they cannot be dominant and underemphasised the attractiveness of pure materials which can never be dominant. E. M. Hoover*[2] *thought that Weber's 'labour coefficient' ought to express a relationship between labour cost and other costs, not weight, in order to estimate its importance.*

Others have said that Weber's ideas on agglomeration were misconceived and do not accept his contention that this phenomenon reinforces the attraction of cheap labour locations. Agglomeration is more likely to be the result of the exploitation of local natural resources, good communications, the availability of large urban markets and a pool of skilled labour.

Weber's industrial location model might seem then to be too divorced from the real world to be valuable. Indeed, one writer considered it to be a positive hindrance to a proper understanding of the timber industry of Oregon. Louis Hamill has written:[3]

A study of industrial locations in Oregon was hindered by Weberian location theory, which interfered with accurate observation.

Also, the information needed to use the theory increased the amount of information to be processed, with no compensating return. Knowledge of location theory often led to erroneous explanations. The theory was unnecessary because industrial locations were understandable without reference to location theory. Attempts to relate the locations of plants to location theory failed because the theory makes unrealistic assumptions about the behaviour of businessmen and about the procedures used in business. It ignores the effects on location of policies and programmes of the government, lending institutions, competing firms, competing industries, etc. It ignores the variety of accounting, financial, organisation, managerial and other practices available to businesses. It erroneously assumes that all locational decisions are made by 'entrepreneurs', whose objective is to maximise profit through the minimisation of costs of assembly, processing and transportation. The efficient practice which the theory assumes for all businesses is found only in unusual cases. The procedures developed by location theorists are not operational; they could not, for example, reproduce the existing pattern of wood-processing plants in Oregon. The procedures actually used in industry are both simpler in broad outline and more complex in operation than those used in location theory. The concept of orientation of industries is inaccurate in view of the great diversity of locational patterns that are observed in every industry.

Notes

1. *The Theory of the Geographical Location of Economic Activities* (Ann Arbor, 1938) p. 19.
2. *Location Theory and the Shoe and Leather Industries* (Cambridge, Mass., 1948) p. 63.
3. 'A Test of Weberian Location Theory', *Abstract of Papers, 20th International Geographical Union* (London, 1964) p. 234.

However, Hamill's views were based on empirical evidence gained from the timber industry in Oregon, and although many of his reservations echo those to which allusion has already been made, there is value in the Weberian model. It validly distinguishes the relative importance of 'ubiquitous' and 'localised' materials, the inputs and outputs to be moved, transport, labour and agglomeration factors. It also differentiates the various industrial orientations, whether they be to materials, to labour or to markets.

THE REFORMED WEBERIAN MODEL

But the model is also capable of some improvement, and as suitable modifications are applied, it will be noticed that the emphasis of Weber's original concept will move from that of 'nearest material source' or 'nearest market' to the concept of 'cheapest material source' and the 'most lucrative market'. In making modifications to the model, it is possible as a start to replace the ton-mileage notion by one which compares value with weight, with changes in weight due to manufacture, with bulk, with fragility and with perishability. (This new concept does in fact sustain Weber's idea about ubiquitous materials since they are abundant and low in value to weight or bulk, and thus costly to move.)

Secondly, all raw materials can be thought of as 'gross' or weight-losing, while 'pure' materials, those that lose none of their weight, can be considered as semi-finished industrial products that are later used to make up a final product.

With these revised considerations in mind, industries will tend to locate:

(*a*) nearer the sources of their materials and fuel
  (*i*) *the greater the amount of waste*
  (*ii*) *the greater the bulk (perishability or fragility)*
  (*iii*) *the higher the freight charges*
  (*iv*) *the lower the value of raw materials compared with products;*
(*b*) nearer the market
  (*i*) *the lower the weight loss*
  (*ii*) *the lower the bulk*
  (*iii*) *the lower the haulage rates*
  (*iv*) *the higher the value of raw materials used.*

Several location possibilities exist for an industry that uses 'pure' materials. It will locate:

(*a*) nearer the source of the major material input *if this has a low value/weight (bulk) ratio;*

(*b*) nearer the labour, the market or at intermediate locations *if these pure materials have a high value/weight ratio and are thus more transportable.*

Industries that assemble several 'pure' materials giving a substantial gain in weight or bulk for the final product (e.g. motor-cars) tend to be market-orientated. The same applies if the industry in question processes divisible materials such as oil. The refining industry tends therefore to be close to its markets.

Turning to the question of markets in greater depth, it is interesting to note that in his book, Weber put forward a notion which forms the basis for further developing his model. He suggested that any economy consists of a series of layers which are related one to another because they represent sources of supply and demand. These layers would form a hierarchy consisting of agriculture and forestry, mining and primary processing, manufacturing and tertiary services.

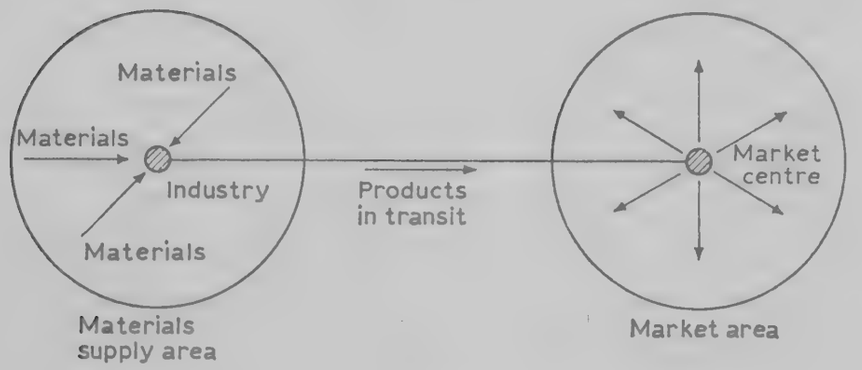


Fig. 20.11

Each layer can be considered as one market area or as a series of market areas for industry as a whole or for individual industries. If a situation can now be envisaged where all the layers are collapsed into one, then what emerges is a complex pattern of centralised markets and resource centres (a coalmine, for example) located in a weaker market field which would reflect low population densities and other non-industrial land uses. One market area for a product would merge with other market areas and with centres or areas which supply materials for the manufacturer. In this respect it has to be remembered that the industrialist will get his materials from a point, several points or a whole area, while distributing his finished products to a point or points, but more usually to one or more areas. All these points and areas will overlap.

Given this situation, the problem of location becomes one of transport costs for the acquiring of materials and the distribution of manufactured goods between points or between point and area or area and point.

When transport costs become decisive in establishing an industry at one point or another, the industry will tend to locate at a point where the areal transport costs for materials, together with transport costs incurred for moving finished products from the manufacturing point to the market centre point and out into an areal market if necessary, are at a minimum. The simplified diagram opposite (Fig. 20.11) helps explain this argument.

It might be thought that the areal supply of materials for manufacturing (as opposed to point supply), or conversely the areal market, could each be dominant features in determining location. However, the 'pull' of market location tends to be more powerful. In the first place, areal product distribution (as opposed to that from the industrial plant to the market central point) tends to magnify the difference between the charges levied for finished goods in the real-world situation and the much lower charges made for moving materials of the same weight and bulk. Secondly, the size of demand in the market area tends to be of vital importance in determining distribution costs, and since the need to maximise sales seems to be more realistic than Weber's original 'least-cost' motive for location, industries are more likely to position themselves to take advantage of this situation. The result may well be a location which gives a large profit but by no means a least-cost *location, since costs of assembly and distribution may be higher than they would be at the least-cost point.*

Finally, it is clear that Weber's model is further improved if his assumptions of a given market and population distribution is supplanted by the recognition of a dynamic situation where areas of industrial development tend to attract new industrial units thus promoting further population and market growth.

However, even though it is possible to build into the Weberian model factors which enhance its utility and bring it more into line with reality, it must not be forgotten that there is still a whole range of other influences which are at work, ranging from government policy right through to the preferences of the industrialist and his own idiosyncratic behaviour. Therefore the final decision to locate at a particular point may be a much more complicated synthesis than even the reformed model can take into account. This notion is supported by work carried out in the U.S.A., since T. E. McMillan has reported on an attempt to find out for a whole range of industries and businesses what are the most important considerations in location.[1] This was done in 1964 through the McGraw-Hill Plant Site Survey in which a questionnaire was sent out to the business and industrial subscribers of a leading United States business journal.

As a result of the replies received, thirty factors or determinants were

identified and could be listed in ranked order according to how each was rated by entrepreneurial respondents. In considering the resulting table (see below), it is apparent that while some of the factors are part of the Weberian model, others are not.

Area of interest	*Percentage of respondents indicating significant interest*
Trucking	76
Reasonable cost of property	67
Reasonable or low taxes	65
Ample area for expansion	63
Favourable labour climate	62
Favourable attitude of community and residents to industry	57
Nearness to present sales area	56
Reasonable cost of construction	53
Favourable climate for personnel	50
Availability of labour skills	48
Access to utilities	43
Near sources of raw materials	42
Need for plant to service new or expanding sales areas	42
Favourable political climate to business	40
Pleasant living conditions	39
Commercial services	37
Rail facilities	35
Zoning restrictions	34
Cost of living and economic conditions	34
Labour rates	33
Educational facilities	32
Favourable climate for productive process	30
Inexpensive fuel and power	29
Public transport	29
Recreational and cultural facilities	25
Water supply	22
Waste disposal	21
Near airport	20
Topography	8
Water transport	8

By way of further explanation of the table and the survey, McMillan has written:

Questions asked by McGraw-Hill were generally ordered [so that] markets were placed first, labour next, etc. Community items such as taxes were placed near the bottom. . . .

The McGraw-Hill survey questionnaire asked a relatively simple question: 'If your company were selecting a new plant site, which of the following considerations would be of importance to you in selecting the specific area or site?' . . . In reply to this question 76 per cent of the respondents mentioned good truck transport. . . . 'Reasonable cost of property' was rated important by 67 per cent of the respondents. This is closely related to 63 per cent which were interested in ample area for expansion. Many industries which have experienced crowding that prevented subsequent expansion at reasonable land prices today wish to purchase land much in excess of the immediate requirements. Highly developed industrial areas with land costs ranging from 15 cents per square foot upward to several dollars would not be likely to meet this requirement. Reasonable or low taxes were mentioned by 65 per cent of the respondents, and a favourable labour climate by 62 per cent.

In declining order the areas of interest rated important by respondents are indicated in the tabulation.

It should be recognised that the percentages given in the tabulation do not measure the importance of individual factors to any particular type of industry. However, they do suggest that for industry in aggregate such items as motor freight services, property costs, taxes and labour climate may deserve more attention than in the past. They also suggest that recreational and cultural facilities and zoning – while rated important by 25 per cent to 34 per cent of the respondents – do not appeal to as many firms as do tax, land costs and labour climate. The writer's experience in industrial development is consistent with the results of the McGraw-Hill survey. Comparative tax data is one of the first areas of interest indicated by site selection teams. The second most frequent area of interest is sites. Proper plant location terminology does not equate land and sites. A site is land of suitable terrain which is served (or is about to be served) by access streets and all utilities of adequate size and capacity for industrial needs. When these sites are removed from congested areas, industry expects reasonable prices. What is reasonable is a function of degree of importance in terms of streets, utilities and terrain, the extent of development and type of use of surrounding parcels of land, present use and alternative uses of the site being considered, and prospects for appreciation in value. Land prices also enter into

considerations of tax reasonableness, since a purchase price of market value will affect future *ad valorem* and franchise tax costs.

Other factors of importance occur with less regularity than do the preceding four factors – taxes, sites, labour and transport. At the same time, these less frequent factors are of prime importance to certain types of industry. A major electronics firm which is interested in research as well as production will show a strong interest in education. From the company's operating standpoint, this interest will be oriented to graduate education for professional employees and to sub-degree programmes for technologist and engineering assistants. Because of the overall high educational level of such firms' employees, an interest will be shown in public school programmes. These will be judged in terms of curricula offered and results of standardised achievement tests.

Industries which have heavy bulky raw materials or products will be particularly interested in rail services. A mild climate will interest firms because it reduces worker absenteeism and may reduce construction costs both in terms of types of buildings and in terms of how much of the operation must be enclosed. Several years ago recreational and cultural facilities were of prime interest to firms which faced massive recruitment of professional personnel; there seems to be a declining interest in this factor. Water supply in terms of volume, quantity and cost is of interest only to the extent that it is adequate for normal sanitary purposes. Public transport has been raised as a factor in very rare instances, usually by firms whose experience has been limited to large metropolitan areas where employees rely heavily on public transport to allow them private conveyance. In the south-west, availability of commercial air services or facilities for commercial aircraft is not an uncommon factor.

While the work reported by McMillan refers particularly to the U.S.A., it is clear that it has the same broad implications for other countries in terms of industrial location, but only provided the differing factors of environment, be they social, political or physical, are taken into account.

Note

1. 'Why Manufacturers Choose Plant Locations *v.* Determinants of Plant Location', *Land Economics*, vol. 41 (3) (1965) 239–46.

URBAN LAND USE

Land use patterns in an urban situation arise from a whole series of individual decisions concerning location. These decisions are controlled in many different ways by the economic processes operating in society.[1] *The relevant characteristics of this mechanism can be summarised briefly as follows. Each urban activity is able to derive utility from all sites in the urban zone. The utility of a site can be measured by the rent the activity is willing to pay. Ultimately competition in the urban land market for the use of available sites results in the occupation of each site by an activity which is able to derive the greatest utility from that site and which is willing to pay most to occupy it. The logical outcome of this concept is the emergence in an urban area of an orderly pattern of land uses in which rents are maximised throughout and all activities are optimally located. This process is the same as von Thünen's formulation of the ordering of land uses in agricultural areas in relation to a market centre. As in von Thünen, in the urban land market we can also explain rent differentials among homogeneous sites by rising (or falling) transport costs.*

Warren R. Seyfried in 'The Centrality of Urban Land Values' (1963)[2] *described the situation as follows:*

Given the assumptions of a competitive market and homogenous sites, accessibility *within* an urban market will be at a maximum at the market centre. As distance from the market centre increases, accessibility decreases if measured in transport and communication costs. Following the models of Alonso, Muth, Wingo and others,[3] market equilibrium results in spatial equilibrium. Firms and households have no incentive to change location, *ceteris paribus*, because profits and other objectives are maximised. Accordingly, output is optimal and the maximum efficiency of the city as a productive unit is realised. The resulting structure of land uses reflects institutional arrangements, e.g. zoning ordinances, the transportation system and policies of financial institutions as well as the competitive ranking of the city, i.e. its economic base. The location of firms and households within this structure depends upon competitive bidding for specific sites. Rent differentials result because market forces require from each site that rent resulting from maximum utilisation or highest and best use. Maximum economic rent occurs at the market centre because the supply of sites and average transportation and communication costs for the local market are least.

Production costs, i.e. wages and interest, are assumed to be independent of location, but transportation costs rise as distance increases so that rent, the surplus of gross revenue over production costs, declines proportionately. The supply of sites, i.e. more and larger land units, increases with distance from the centre. Competing users of sites will locate relative to their economic rent potential so that a structure of site values relative to location results from market forces. This structure of the urban land market can be visualised as a rent or land value surface; the market centre is the apex, the point or area of highest site value. As rent declines with distance, value declines and land uses change. If sites of equal value are connected, the iso-value lines are contours of a cone; the differential between the iso-value lines or contours is a function of site rent or value. Thus, the spatial structure of land uses or the urban land market at a moment of time can be shown by the rent or value function or gradient.

For a metropolitan area having a broad economic base, the rent surface would reveal a hierarchy of cones, nodes or apexes resulting from nucleation of competing economic functions. The more accessible the site, the greater the nucleation, and the steeper the slope of the value gradient. The area of intersection of rent functions resulting from nodes within the rent structure would show thresholds of economic dominance.

Notes

1. Except in cases where local authorities or governments predetermine the use to which a unit of land may be put.

2. *Land Economics*, vol. 39, pp. 275–6.

3. Differentiation among sites because of quality factors is also a determinant of economic rent. In this sense, one recalls Ricardian rent theory: Rent differentials arise because of differences in use capacity. Thus, urban sites vary in rent and value because of use capacity as well as location. This article considers only location or position rent resulting from accessibility.

THE LAND-VALUE SURFACE

A visualisation of the generalised land-value surface within a city which Seyfried described above is to be found in 'Commercial Structure and Commercial Blight' (1963) by B. J. L. Berry and others.[1] *It is reproduced below.*

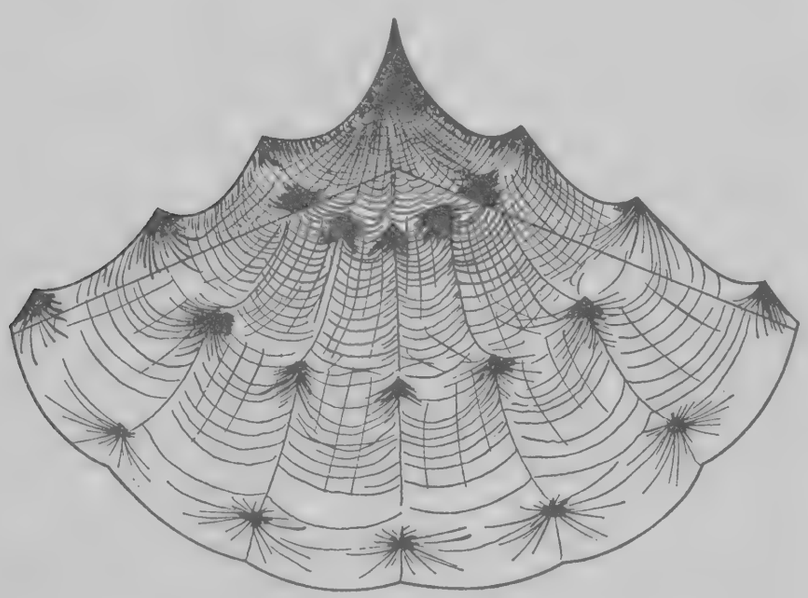


Fig. 21.1 General pattern of urban land values

Note

1. Department of Geography, University of Chicago, Research Paper No. 85 by B. J. L. Berry, R. J. Tennant, B. J. Garner and J. W. Simmons, p. 14.

The land-value surface is an important concept because it can be looked upon as the common denominator of urban land uses. An appreciation of the pattern of land values is therefore a substantial aid to the understanding of the internal structure of cities. Any pattern of land values in a given city, just like its equivalent pattern of land uses, will be unique since local circumstances will never be exactly repeated elsewhere. But we can argue that the land-value surface is essentially a direct reflection of accessibility within the urban area. This will be greatest at the city centre because it has developed over time as the focus of routes and is therefore the most accessible part of the whole urban zone. Competition for land use is most intense there. Similarly, accessibility is greater at sites located along the radial and circumferential routes and at the intersections than away from them. As some parts of the urban area will have better transport facilities than others, the value surface will contain distortions of slope, etc., and will not remain uniform in all directions.

It should be stressed that the term accessibility will vary in its implications for different activities. In industry for instance it must relate directly to costs of operation and profit levels, but for residential activities proximity to work, shopping centres and other environmental intangibles will be important.

The concept of the land-value surface as it has been identified has found support in several empirical studies beginning with that of H. Hoyt in Chicago (1939). In his book 'One Hundred Years of Land Values'[1] *he wrote:*

If the land values in Chicago were shown in the form of a relief map, in which the elevations represented high land value, a picture of startling contrasts would be disclosed. In the centre would be the Himalaya Mountain peaks of the Loop, but on all sides except along a high ridge running north along the lake there would be a descent into the deep valleys of the blighted areas. Gradually, as one went farther from the centre, the elevation would begin to rise. Along the lake, both north and south, would be a high ridge which slopes down sharply as one goes west. Beginning five or six miles from the centre of the city, there would be a plateau several miles wide encircling the city that is uptilted towards the lake, on top of which were high ridges a mile apart that culminated in towering pinnacles at each intersection. If a dollar a front-foot in land value were represented by 1 ft in height on this map, the changes in elevation within the 211 square miles of Chicago's area would be greater than the differences in physical elevation of any part of the land surface of the globe, for the variation would be the same as from 5 to 50,000 ft above sea-level. Within a little over a mile one might drop from an altitude of 50,000 ft to one of 50 ft,[2] and within a short block a person could fall from an elevation of 4000 ft to one of 25 ft.[3]

Notes

1. University of Chicago, pp. 297–8.
2. From State and Madison Street to Townsend Street near Chicago Avenue.
3. From Forty-seventh and Ashland Avenue to Forty-sixth and Laflin.

More recently, the most comprehensive study of the land-surface-value concept yet undertaken was made by Maurice H. Yeates in Chicago. Working from the basic assumption that land values decline with distance rom the centre of the city, he stated in his paper 'Some Factors Affecting the Spatial Distribution of Chicago Land Values, 1910–1960' (1965)[1] that:

In this study the centre of the city is defined as the peak-land-value intersection of State Street and Madison Avenue. This point, in Chicago, is close to the geographic centre of the central business district, which is commonly defined as that area surrounded by the rapid-transit elevated railroad, known locally as the 'Loop'. The peak-value intersection has arisen because transport routes converge at the centre of the city making that location the point of minimum aggregate travel costs. Locations away from the centre incur greater transport costs, and thus land values decline reflecting decreasing accessibility, smaller market hinterlands and lower net returns.

However, the central business district is but one, although it is the largest, of a number of commercial areas in the city. If land values rise towards the peak-value intersection of the central business district, they should also rise towards the highest-value intersections of the smaller commercial areas. Recognising this, both Knos and Marble[2] attempted to relate land values to distance from major business centres as well as the central business district. The evidence from these studies is somewhat contradictory, for in the former the assumptions were substantiated and in the latter rejected. As a consequence these ideas will be re-examined in the Chicago area, at decennial intervals since 1910, in relation to the central business district (the Loop) and regional shopping centres.

The city, however, presents an ever-changing pattern of accessibility. These varying patterns are a result of physical changes in the structure of internal linkages, such as the construction or widening of highways, and technological changes in transport media, such as the development of faster and more efficient rapid transit, or the introduction of small, low-cost automobiles. Physical improvements, in particular, tend to make one area more accessible, and therefore more desirable, than another that does not have the same facilities. This is because increases in accessibility tend to reduce transport costs, and therefore not only provide additional funds that can be used for the purchase of land but also increase the number of potential users and range of use for that particular area. Evidence from the Chicago suburbs, for example, indicates that there are relatively lower land values in areas without commuter rail service or at some distance from such a service.

Since the turn of the present century the City of Chicago has witnessed the extension and decline of such a physical transport improvement. The elevated-subway system reached its maximum coverage in 1930 and has subsequently decreased in total mileage. This is a reflection of the primacy of public transport in the first three decades of the century, and its replacement by the automobile during the last three decades. Thus one would expect the physical location of rapid-transit facilities to reach its maximum importance as a determinant of land values in 1930 and to decline thereafter.

But this decline in physical extent and coverage of rapid transit cannot be divorced from social and economic changes in society that have occurred during the period under discussion. Just as it is everywhere evident that the automobile is replacing public transit as the chief mover of people in the city, so it is becoming increasingly evident that people as such are changing their location preferences. These changes in location preferences are evidenced by the general decline of population densities in the city and the spreading out of population beyond the political limits of the city into the suburbs and interstitial areas. The factors that have permitted this movement are many and interactive. Not the least of them are: the universal use of the automobile, the construction of highways and expressways, the shorter working week, increasing wages and increased leisure.

The net impact of such changes as far as land values are concerned should appear to be a more rapid growth of land values at the periphery of the city and in areas offering certain amenity advantages. Thus, throughout the period under discussion, the postulated negative relationship between land values and distance

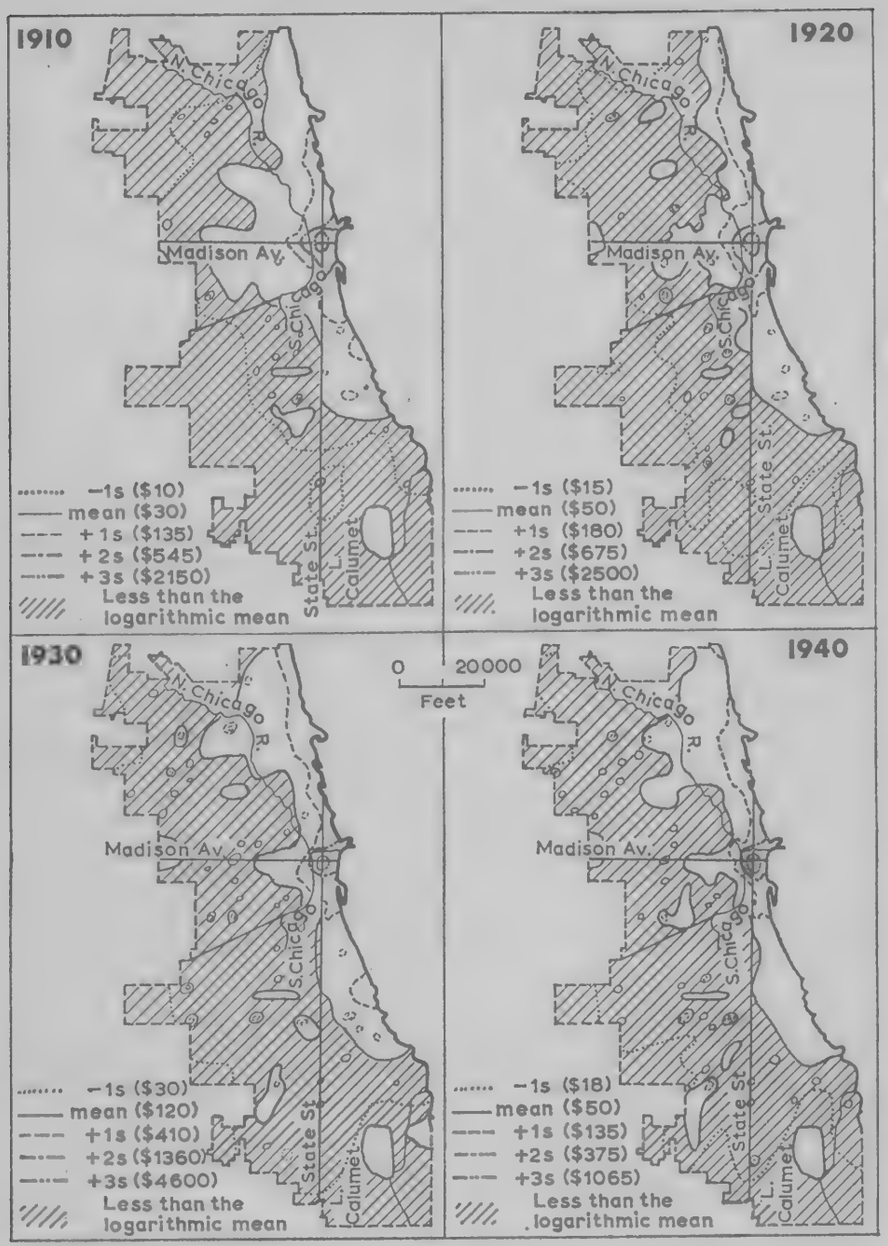


Fig. 21.2 The transformed land value maps, Chicago, 1910–40
Source: *Olcott's Land Values Blue Book of Chicago*, 1911, 1921, 1931, 1941. Each book refers to the situation in the previous year.

from the centre of the city should weaken. This situation may be modified in the case of Chicago because of the truncated nature of the city. The central business district is located adjacent to Lake Michigan and the city has grown to the north and the south. The lake shore today exhibits the geographic juxtaposition of a number of factors, such as the Outer Drive (a high-speed limited-access highway), and part of the elevated-subway system, as well as being an amenity and prestige location. All these factors would seem to suggest that land values should decline with distance from Lake Michigan.

Changing location preferences can be further examined by analysing the relationship between land values and population densities. If there is a negative relationship between land values and distance from the centre of the city, and if population densities decrease with distance from the centre of the city, then land values and population densities should be positively related. However, this relationship may be becoming more and more affected by changing characteristics of the urban population. For example, the proportion of the city classified as non-white (which in reality means Negro) has increased enormously during the past few decades. In 1940, 8·3 per cent of the total population of the City of Chicago was classified as non-white. By 1950 this proportion had increased to 14·1 per cent and by 1960 reached 24·1 per cent. The highly clustered nature of the distribution of these non-white persons makes the net effect of this movement difficult to determine.

LAND VALUE MAPS

The land value maps (Figs 21.2 and 21.3) were constructed from a disproportionate stratified random sample of 484 points (Table 21.1), for which average front-foot land values could be obtained on a block basis, located within the 1960 city limits of Chicago. Preliminary analysis of the arrayed land-value data suggested a sharp positive skewness at each time period. In order to facilitate later statistical analysis the data were logarithmically transformed, and in order to facilitate comparison of the land-value maps between time periods, the transformed data were converted into normal standard deviates. The isoline interval chosen for each map is one standard deviation; therefore each map can be compared because the intervals are comparable and related to their respective means.

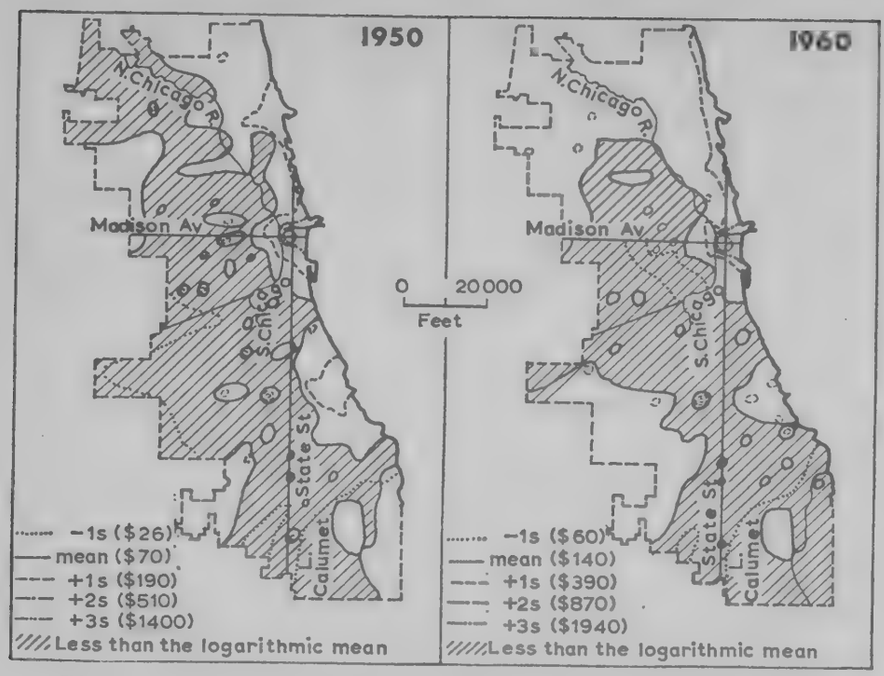


Fig. 21.3 The transformed land value maps, Chicago, 1950–60.
Source: as Fig. 21.2

A visual analysis of Figs 21.2 and 21.3 yields some interesting observations. Throughout the whole time period the highest land values are found within and adjacent to the central business district (the centre being the peak-value intersection at State and Madison), and from this area there is a sharp decline in all directions. This decline was logarithmically fairly uniform in 1910, but by 1960 it does not exhibit the same consistency. The contraction of the mean isoline and general reduction of land values on the south side of the central business district is marked by the presence of residual outliers which indicate the position of regional business and shopping centres. Therefore, it could be suggested that the effect of distance from, and consequently access to, the central business district on land values appears to have diminished in importance during the past fifty years, while the front-foot value of locations in the vicinity of outlying regional service centres has increased.

TABLE 21.1

SAMPLING RATIOS AND NUMBER OF POINTS SAMPLED BY ZONES

*Zone**	*Percentage of total area in city*	*Sampling ratio*	*Number of sample points*
A(0 –1·5 miles)	2	3:1	30
B(1·5–3·0 miles)	7	2:1	70
C(3·0–4·5 miles)	10	12:10	60
D(rest of city)	81	8:10	324

* The zones are rings centred on the peak-value intersection of State Street and Madison Avenue, Chicago.

A further locational factor that appears to influence the general spatial distribution of land values is Lake Michigan, for relatively higher land values are found close to the lake shore. The 1950 and 1960 maps, however, suggest that land values within the vicinity of Lake Michigan on the north side of the city are generally increasing and higher than those on the south side where values are declining. This is possibly a reflection of the non-white, low-income nature of the southern area which discourages 'high rise' apartment speculation.

Land at the periphery of the city appears to be becoming relatively more valuable than land in the middle of the city.[3] In the 1960 map this tendency is much more strongly developed in the north than in the south-west, and hardly present at all in the west. The great rise in land values at the periphery of the city in the north is probably a reflection of the high-income nature of this area. The absence of any rise to the west is a result of the political boundary on the west being closer to the centre of the city than the boundary to the north or south, and thus peripheral developments which are within the city on the north or south are absent in that area.

Notes

1. *Economic Geography*, vol. 41 (1965), pp. 57–62, 67–8.
2. D. S. Knos, 'Distribution of Land Values in Topeka, Kansas' (University of Kansas, 1962); D. F. Marble *et al.*, *Studies of Highway Development* (Seattle, 1959).
3. The descriptive term 'middle of the city' refers to the area between the peripheral zone and the Central Business District, and the fringe (Zone A in this analysis).

Having established the nature of the land-value surface, Yeates then sought to explain the variation in the value surface for Chicago by the use of a multiple regression model.[1] *In this he attempted to handle six variables which were as follows:*

1. *Distance from the central business district* C'_i
2. *Distance from the nearest regional shopping centre* R'_i
3. *Distance from Lake Michigan* M'_i
4. *Distance from the nearest elevated-subway system* E'_i
5. *Percentage non-white population of the block in which a particular site is located* N'_i
6. *Population density* P'_i

The model was examined decennially for the years 1910–60 using the Chicago data. Having established his model, Yeates then wrote:[2]

The application of the above hypothesised model at each one of the time periods yields some interesting results. In 1910 and 1920 the model fits the actual structure of land values extremely well, but in 1930 the explanation drops appreciably, and by 1950 and 1960 the 'explanation' is extremely low (Table 21.2). This may well be a reflection of the decline in importance of a number of variables in the model. For example, an examination of Table 21.3 suggests that the regression coefficient pertaining to the variable distance from the central business district has declined in magnitude considerably throughout the whole time period, even though it tests as being significantly different from zero in 1960. The regression coefficient pertaining to the variable distance from the nearest elevated-subway system fails, however, to test as being significantly different from zero in 1960. These changes in magnitude and statistical significance may well be a result of actual changes through time in how these variables affect land values, or the result of inadequate transformations chosen for the individual variables, which for all variables, except percentage non-white, has been presumed to be logarithmic. The distribution of the individual variables was tested at each time period and a logarithmic transformation appeared to be most appropriate to assure normality.

Thus is would seem tenable that the decline in coefficient of determination through time may well be a result of the decline in importance of certain variables. One way of investigating this would be to make a comparison of the regression coefficient (or slope) between each independent variable (all other independent variables held constant) and the dependent variable. In this particular case the regression coefficients are not comparable because the independent variables are expressed in different units. Comparability can be achieved, however, if the dependent and

TABLE 21.2

RESULTS OF MULTIPLE CORRELATION ANALYSIS, 1910–60

	1910	*1920*	*1930*	*1940*	*1950*	*1960*
Multiple correlation coefficient (R)	0·8784	0·8070	0·6089	0·5846	0·4885	0·4216
Coefficient of determination (R^2)	77%	65%	37%	34%	24%	18%

TABLE 21.3

CONSTANTS AND REGRESSION COEFFICIENTS FOR MULTIPLE REGRESSION ANALYSIS, 1910–60

	a	*C'*	*R'*	*M'*	*E'*	*P'*	*N*
1910	4·745	− 0·837[a]	− 0·038	− 0·450[a]	− 0·248[a]	+ 0·105[a]	+ 0·005[a]
1920	4·826	− 0·673[a]	− 0·122[a]	− 0·414[a]	− 0·240[a]	− 0·008	+ 0·001
1930	4·107	− 0·268[a]	− 0·156[a]	− 0·367[a]	− 0·214[a]	+ 0·039	− 0·003[a]
1940	3·387	− 0·275[a]	− 0·134[a]	− 0·285[a]	− 0·140[a]	+ 0·044	− 0·002[a]
1950	3·562	− 0·268[a]	− 0·081	− 0·227[a]	− 0·152[a]	− 0·116[a]	− 0·002[a]
1960	3·356	− 0·173[a]	− 0·092[a]	− 0·146[a]	− 0·050	− 0·137[a]	− 0·002[a]

[a] Significantly different from zero at $t(0·05)$. For a discussion of the computation and use of t, see H. M. Walker and J. Lev, *Statistical Inference* (New York, 1953) pp. 337–9.

independent variables are expressed in standardised form. The regression coefficients are then called standard regression coefficients.

The regression coefficients have been standardised for each year, and on the basis of comparison of slopes a ranking can be made of the importance of each variable at each time period in this particular model. A comparison of these rankings (Table 21.4) indicates that the distance measures have been particularly important, especially those related to the central business district, the elevated-subway system and Lake Michigan. All these are distance variables; they measure airline distance to important facilities, such as transport and recreational amenities. Of these the distance measure concerning the central business district has been interesting in that the decline in importance of this variable has been largely responsible for the great decline in power of the model. The sign of the coefficient indicates that on the whole land values do decline with distance from the central business district, but the standardised slope of this variable has declined by some 60 per cent in the last forty years, illustrating a decline in attractive power of the central business district. Not only has there been a decline in absolute magnitude of the central business district standard regression coefficient, but also a decline in its relative position among the other variables.

Along with the decline in importance of proximity to the central business district has been the marked decline in the ranking and absolute magnitude of the standard regression coefficient related to distance from the nearest elevated-subway. The theoretical discussion suggested that this variable should reach maximum importance in 1930 and decline thereafter. The rankings indicate that this was only partially true (Table 21.4), for this variable ranked second in 1930, but first in 1950. The tendency for a decline after 1940 was, perhaps, delayed by the special circumstances of the 1940–50 period which emphasised the importance of mass transit. This recent great decline may have been accentuated by the great increase in peripheral employment opportunities and the recent construction of Loop-oriented expressways. Hoch used Olcott's land values to examine the relationship between new expressway facilities and the value of adjoining land, taking the case of Edens Expressway from Foster Avenue to just north of Lake-Cook Road.[3] Only a portion of the survey was within the City of Chicago, and although land values tended to rise towards Edens, the lowest increases were recorded in that portion of the study in the city. Thus it would appear that although transport improvements result in an increase in land values, the greatest increases are occurring beyond the city boundaries, and those within the city hardly result in increases at all.

The great rise, both relative and absolute, in importance of the standard regression coefficient related to the non-white variable has occurred during the past decade. The interpretation, for the city as a whole, must be that as the percentage non-white in an area increases, land values decrease. It must be remembered, however, that this variable is quite closely associated with income. The statistical analysis, therefore, supports the map analysis which suggested that lower land values in the City of Chicago were generally associated with low-income non-white areas.

In the multiple regression model the influence of Lake Michigan on the spatial distribution of land values also supported the conclusions of the map analysis. The regression coefficient pertaining to this variable tested as being significantly different from zero throughout the whole time period and the sign was, as expected, negative (Table 21.3). Furthermore, the table of rankings (Table 21.4) indicates that this variable has maintained a high relative importance throughout the whole time period. This supports the presumption that the geographic juxtaposition of the Outer Drive, part of the elevated-subway system and the amenity and prestige advantages of the lake shore have all combined to maintain the important influence of this variable on land values.

TABLE 21.4

COMPARISON OF RANKING OF STANDARD REGRESSION COEFFICIENTS

Rank	*1910*	*1920*	*1930*	*1940*	*1950*	*1960*
1	*C′*	*C′*	*M′*	*M′*	*E′*	*N*
2	*M′*	*M′*	*E′*	*C′*	*M′*	*M′*
3	*E′*	*E′*	*C′*	*N*	*C′*	*C′*
4	*P′*	*R′*	*N*	*E′*	*N*	*P′*
5	*N*	*N*	*R′*	*R′*	*P′*	*R′*
6	*R′*	*P′*	*P′*	*P′*	*R′*	*E′*

The general influence of regional centres did not, however, prove to be as important as the visual analysis of the land-value maps suggested. Although the regression coefficient tested as being

significantly different from zero in four out of the six time periods, and the sign supported the hypotheses that land values decline with distance from regional centres, the value of the standard regression coefficient ranked very low among the other variables. A reason for this may well be that the influence of regional centres on land values is spatially extremely local, resulting in a very tight peaking of land values. Also, the regional centres over the whole city exhibit a variation in stage of growth. It could be expected that relatively higher values would be found in regional centres of recent growth and lower values in the older and declining centres.

The interpretation of the results obtained from the analysis of the relationship between land values and population density is rather more complex than the initial theoretical discussion suggested. It will be recalled that the theoretical discussion suggested that 'if there is a negative relationship between land values and distance from the centre of the city, and if population densities decrease with distance from the centre of the city, then land values and population densities should be positively related'. The analysis, so far, has indicated that land values and distance from the central business district are inversely related, but that this relationship has declined during the past forty years. With this modification one would expect the hypothesised positive relationship between land values and population density to weaken through time. Examining the second part of the argument, it appears that, over the whole city, there has been a distinct decline in the negative relationship between population density and distance from the central business district (Table 21.5) which would further emphasise the decline in the positive relationship between land values and population density.

TABLE 21.5

SIMPLE CORRELATION COEFFICIENTS (R): POPULATION DENSITY (P') AND DISTANCE FROM THE CENTRAL BUSINESS DISTRICT (C')

	1910	*1920*	*1930*	*1940*	*1950*	*1960*
R	− 0·63	− 0·56	− 0·23	− 0·18	− 0·20	+ 0·04

The multiple regression analysis suports this hypothesis modification. In 1910 the regression coefficient pertaining to this variable supported the hypothesis of a positive relationship. For the next three time periods the regression coefficient did not test as being significantly different from zero, and in 1950 and 1960 there was a statistically significant negative relationship. That is, over the whole city, land values increased as population density decreased. The reasons for this are threefold: first, the general flattening of the population density curve through time; second, declining central densities and the outward movement of the inflection point of the population density curve (which is the result of a number of sample points located in the commercial centre of the city having both very high land values and very low population densities); and third, increasing land values at the periphery of the city.

Notes

1. In geographical research regression models are used to analyse the degree and direction of correspondence among two or more spatial patterns or locational arrangements. Regression model analysis can be guided by hypotheses about functional relationships or causal mechanisms between the phenomena under discussion, or they can be exploratory – trying to derive inductive generalisations about the covariation of spatially distributed phenomena.
2. *Economic Geography*, vol. 41 (1965) 62–6.
3. I. Hoch, 'Rising Land Values found along Edens', *Cook County Highways*, vol. 4 (1957) 4.

Put simply, Yeates's argument set out above states that the model provides an excellent degree of explanation of the variation in the land-value surface when applied to the data taken from early decennial periods. For example, in 1910 it gives a 77 per cent explanation (Table 21.2). However, the model provides a low degree of explanation of the variation in the surface, only 18 per cent when applied to data for 1960 (Table 21.2).

But in his visual analysis of the land value maps, Yeates did notice broad areas of similar land values that suggested the drawing-up of a series of regions or sectors.[1] *(We shall refer to these again later.) Yeates went on to say that these regions 'can be built into the regression equation by the use of dummy and interaction variables so that the (original) regression equation is enlarged'.*[2] *These additional variables were nominal variables representing his delineation of a North sector, a North-west sector, a West sector, a South-west sector and a South sector (see Fig. 21.8 above). He then added*[3] *that:*

Solution of the equation for the 1960 data resulted in an estimate of the parameters for the variables within each of the six sectors, and a coefficient of determination of 51 per cent. Thus the hypo-

thesised model, which was expanded to incorporate the concept of 'within-city' sectoral variations, resulted in an increase in 'explanation' of about 33 per cent. Therefore, it could be suggested that given the variables and the basic model, sectoral variations, arising out of differing stages of historical growth and differing spatial characteristics, are beginning to account for a highly significant proportion of the variation of urban land values in Chicago.

An examination of the signs of the regression coefficients (Table 21.6) provides support for a number of conclusions reached in the

TABLE 21.6

COMPARISON OF SIGN OF REGRESSION COEFFICIENTS BY SECTOR, 1960

Sector	*C'*	*R'*	*M'*	*E'*	*P'*	*N'*
North	$+^a$	−	$-^a$	−	−	−
North-west	$+^a$	$-^a$	−	$-^a$	+	+
West	−	$+^a$	−	−	−	$-^a$
South-west	+	−	$+^a$	−	+	$-^a$
South	$+^a$	−	$-^a$	$-^a$	$+^a$	−
Zone A	$-^a$	−	−	−	+	+ −
Whole city	$-^a$	$-^a$	$-^a$	−	$-^a$	$-^a$

[a] Significantly different from zero at t (0·05).

previous section, and illuminates the relationship between land values and population densities. In only one sector in 1960 was the population density coefficient significantly different from zero and that was in the South sector, where there did appear to be a positive relationship. Furthermore, examination of the population density and per cent non-white variables in the West and South-west sectors reveals that land values and percentage non-white are negatively related (conformant with the theoretical concepts), but that population density is insignificant. However, the South sector, which has been the recipient of Negro urban immigration since the turn of the century, shows no significant decline in land values as the per cent non-white in an area increases. This may be related to the 'piling-up' process mentioned by Duncan.[4] In areas of recent non-white incursion and establishment land values drop, but as the number of people increase in the neighbourhood land values may become stable and increase as the population density increases. The area of greatest 'piling-up' is in the South sector; and the idea is given some credence by the fact that this was the only sector where there was a significant positive relationship between land values and population density.

As far as the other variables are concerned, this analysis by sector suggests that although the influence of the elevated-subway may be declining over the whole city, it still has a significant effect on land values in two sectors. Lake Michigan influences, as would be expected, those sectors adjacent to it. The positive relationship in the South-west can be dismissed, as here the variable has little meaning except as a measure of proximity to the periphery. Land values rise towards regional shopping centres in the newer North-west, but decline in the vicinity of such centres in the older and more industrialised West.

The regression coefficients substantiate quite adequately conclusions already reached concerning the influence of proximity to the central business district. In three sectors there proved to be a significant rise in land values with distance from the centre of the city. Only in Zone A (demarcating an area within 1·5 miles of the central business district) did land values decline with distance from the intersection of State and Madison.

Notes

1. See *Economic Geography*, vol. 41 (1965) 62.
2. Ibid., p. 67
3. Ibid., pp. 67–8.
4. O. D. Duncan and B. Duncan, *The Negro Population of Chicago* (Chicago, 1957).

Yeates, in summarising the main points of his complete thesis, concluded:[1]

A mathematical model for 1910, expressed in multiple regression format, 'explained' 77 per cent of the variation in Chicago land values. However, there was a continuous decline in 'explanation' derived from a multiple regression model of the same form applied during ensuing time periods. For 1960 this 'explanation' had dropped to 18 per cent. This decline in power of the model can be related in particular to the decline in influence of the central business district and the increase in importance of sectoral variations. Addition of sectors raised the 'explanation' in 1960 to 51 per cent, which seems to suggest that sectoral differences at this period accounted for a further 33 per cent of the variation in land values.

Note
1. *Economic Geography*, vol. 41 (1965) 67–8.

There is no doubt that the level of explanation provided by the sectors attest the importance of radial routes in determining land values.

Land use models

So far the nature of urban land use has been examined only in terms of the land surface value. However, it has already been noted in the Yeates paper that the analysis of land-surface-value data can lead to the division of the urban area into well-defined sectors. An analysis of other land use data which must be an expression of the land-value surface can therefore also lead to ordered expression of land use patterns within cities. These tend to conform to models which have been identified under three headings: concentric, sector and nuclei models. Each of these reflects well-defined aspects of the land-value-surface concept. The concentric model is, for example, developed on the assumption that land values, and by implication accessibility, decline with equal regularity in all directions. As distortions which would be caused by differential accessibility are disregarded, land use patterns are assumed to take the form of a series of concentric zones about the city centre, as in the top left of Fig. 21.4 above:[1]

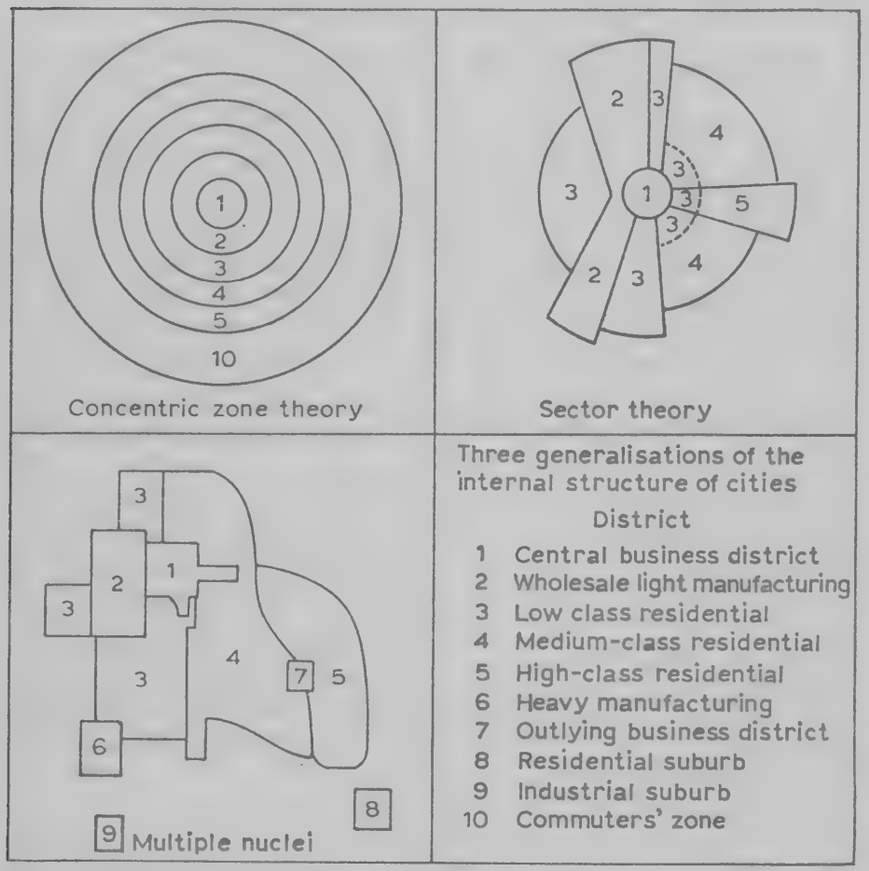


Fig. 21.4 Generalisations of internal structure of cities. The concentric zone theory is a generalisation for all cities. The arrangement of the sectors in the sector theory varies from city to city. The diagram for multiple nuclei represents one possible pattern among innumerable variations.

Note
1. Diagram taken from C. D. Harris and E. L. Ullman, 'The Nature of Cities', *Annals of the American Academy of Political and Social Sciences*, vol. 242 (1945) 17.

The classic concentric zone model was proposed by E. W. Burgess in 'The City' (1925),[1] *and was based on his studies of Chicago.*
He wrote:

The typical processes of the expansion of the city can best be illustrated, perhaps, by a series of concentric circles (see Fig. 21.5), which may be numbered to designate both the successive zones of urban extension and the types of areas differentiated in the process of expansion.

This chart represents (on the right-hand side of the interpolated shoreline) an ideal construction of the tendencies of any town or city to expand radially from its central business district – on the map 'The Loop' (I). Encircling the downtown area there is normally an area in transition, which is being invaded by business and light manufacture (II). A third area (III) is inhabited by the workers in industries who have escaped from the area of deterioration (II) but who desire to live within easy access of their work. Beyond this zone is the 'residential area' (IV) of high-class apartment buildings or of exclusive 'restricted' districts of single family dwellings. Still farther, out beyond the city limits, is the commuters' zone – suburban areas, or satellite cities – within a thirty- to sixty-minute ride of the central business district.

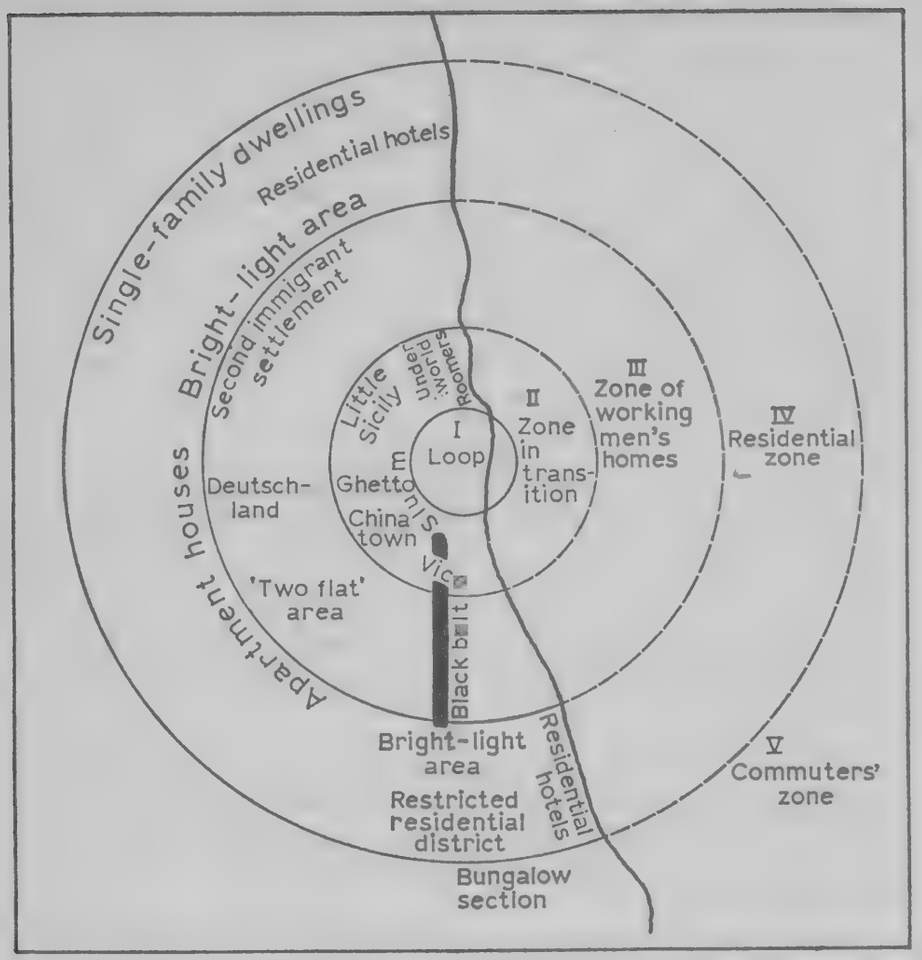


Fig. 21.5 Urban areas

This chart brings out clearly the main fact of expansion, namely, the tendency of each inner zone to extend its area by the invasion of the next outer zone. This aspect of expansion may be called *succession*, a process which has been studied in detail in plant ecology. If this chart is applied to Chicago, all four of these zones were in its early history included in the circumference of the inner zone, the present business district. The present boundaries of the area of deterioration were not many years ago those of the zone now inhabited by independent wage-earners, and within the memories of thousands of Chicagoans contained the residences of the 'best families'. It hardly needs to be added that neither Chicago nor any other city fits perfectly into this ideal scheme.

Note

1. Chap. 2, pp. 50–1, 54–6.

He then considered the case of Chicago (shown on the left of Fig. 21.5 beyond the Lake Michigan shoreline) in greater detail so as to modify his initial concept:

In the expansion of the city a process of distribution takes place which sifts and sorts and relocates individuals and groups by residence and occupation. The resulting differentiation of the cosmopolitan American city into areas is typically all from one pattern, with only interesting minor modifications. Within the central business district or on an adjoining street is the 'main stem' of 'hobohemia', the teeming Rialto of the homeless migratory man of the Middle West. In the zone of deterioration encircling the central business section are always to be found the so-called 'slums' and 'bad lands', with their submerged regions of poverty, degradation and disease, and their underworlds of crime and vice. Within a deteriorating area are rooming-house districts, the purgatory of 'lost souls'. Nearby is the Latin Quarter, where creative and rebellious spirits resort. The slums are also crowded to overflowing with immigrant colonies – the Ghetto, Little Sicily, Greek-town, Chinatown – fascinatingly combining old-world heritages and American adaptations. Wedging out from here is the Black Belt, with its free and disorderly life. The area of deterioration, while essentially one of decay, of stationary or declining population, is also one of regeneration, as witness the mission, the settlement, the artists' colony, radical centres – all obsessed with the vision of a new and better world.

The next zone is also inhabited predominatingly by factory and shop workers, but skilled and thrifty. This is an area of second immigrant settlement, generally of the second generation. It is the region of escape from the slum, the *Deutschland* of the aspiring Ghetto family. For *Deutschland* (literally 'Germany') is the name given, half in envy, half in derision, to that region beyond the Ghetto where successful neighbours appear to be imitating German

Jewish standards of living. But the inhabitant of this area in turn looks to the 'Promised Land' beyond, to its residential hotels, its apartment-house region, its 'satellite loops' and its 'bright light' areas.

This differentiation into natural economic and cultural groupings gives form and character to the city.

Burgess does not attempt an explanation of his concentric pattern, and for this reference has to be made to B. J. L. Berry in 'Studies of Highway Development and Geographical Change' (1959).[1] *In this it is assumed that accessibility falls uniformly from the city centre in all directions and that the rent an activity will pay for a site decreases from the central point. In Berry's diagram (Fig. 21.6) rent curves will therefore slope to the right as in activity* X_t.

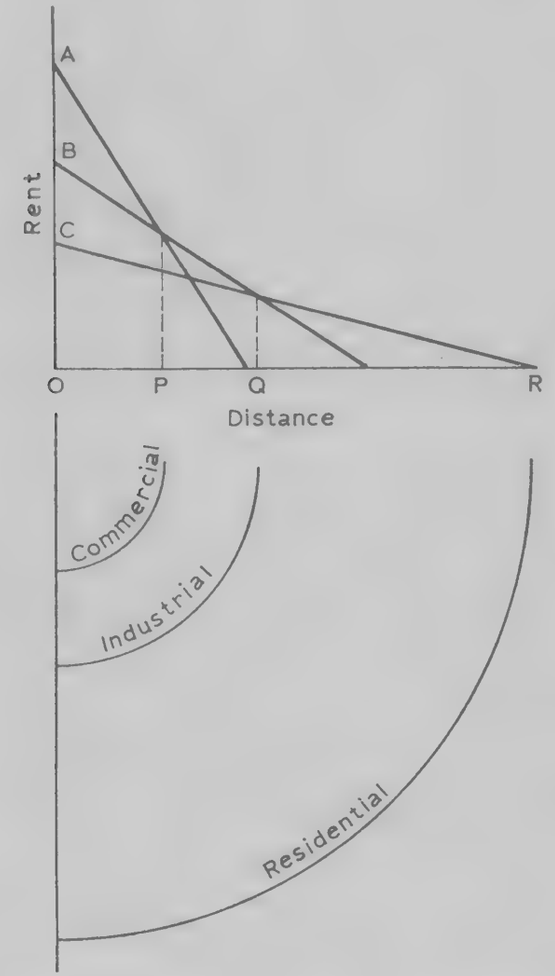


Fig. 21.6 Hypothetical rent–distance relationships within a city

Note

1. By B. J. L. Berry, W. L. Garrison, D. F. Marble, J. D. Nystruen and R. C. Morrill, p. 64.

As the susceptibility of various activities to variations in accessibility differ, their rent curves will vary in steepness, as, for example, in activities X_t *and* X_n. *As the rent curves for* X_t, X_t X_n *are ranked in steepness so they are ranked in order of accessibility from the urban centre. When competition for sites occurs,* X_t *outbids all other potential users and occupies sites from the urban centre O to distance A.* X_t *is then able to pay the highest rent and occupies sites from A to the point on its right where the second broken vertical line meets the horizontal axis.* X_n *can only bid for sites between the last-mentioned point and C. As can be imagined from the diagram, this results in a concentric pattern of land uses and in this case* X_t, X_t *and* X_n *can be respectively thought of as commercial, industrial and residential uses.*

A striking curiosity of the concentric model might seem to be the fact that poor families live close to the urban centre on high-value land while the rich live at the city edge where land is cheaper. Some writers[1] *have explained this by stating that each household has a limited income which must satisfy the basic costs of living, housing and transport to and from work in the city centre. If basic living costs are accounted for, the rest of income can be used for housing and travel and allocated in varying amounts between each. Poor families with less money for travel live close to their jobs but can only afford to pay for small quantities of the high-value land close to the urban centre. On the other hand, rich families are able to luxuriate at the edge of the urban area on large amounts of lower-value land at the expense of a heavy outlay in travel costs to and from the city centre.*

Since Burgess proposed his model, others have identified the concentric pattern model. George W. Hartmann in 'The Central Business District: A Study in Urban Geography' (1950)[2] *wrote as follows:*

The concentric zone hypothesis as outlined by Burgess suggests that, in the absence of topographical or man-made barriers, the geometric pattern is a circle. This belief appears to be based upon

measurement of relative distances from the centre: each point on the periphery of a circle is as close to the centre as any other point similarly situated, and with the desire for central position on the part of commercial activities the entire business district assumes a circular shape around the central point.

It might be argued that developing idealised shapes on the basis of distance measurements alone is of questionable value in that the importance of relative time involved in travel from one point to other points is ignored. That time and distance factors are not synonymous in reality hardly needs demonstrating here.

Actual movements of goods and people, however, ordinarily cannot follow the shortest airline paths to and from the centre. To accomplish this end, it would be necessary for every portion of the area to be a part of some straight-line route to the centre, an impracticable and impossible feat because of the presence of buildings and other space occupiers. Only the streets and sidewalks may be used. The circular pattern is most nearly approached in actuality where the major transportation routes are of the radial type, and then only if the number of radial streets is very large. The periphery of the central district reaches its greatest outward extent along each of the radial streets, and consequent inward sagging occurs in the interstitial areas. The width of these interstitial zones varies inversely with the number of routes to the centre; an actual circular shape requires innumerable radial streets.

As the number of radial thoroughfares decreases, however, the peripheral configuration becomes less circular and the areas where sagging occurs become larger and larger. A small number of such streets imposes a pattern which is more star-like in shape. It may be noted, therefore, that circular and star-like patterns are similar in nature. They differ only in the number of radial thoroughfares assumed.

Notes

1. W. Alonso, 'A Theory of the Urban Land Market', *Regional Science Association, Papers and Proceedings*, vol. 6 (1964) 149–57; L. Wingo, *Transportation and Urban Land* (1961).
2. *Economic Geography*, vol. 26 (1950) 238–9.

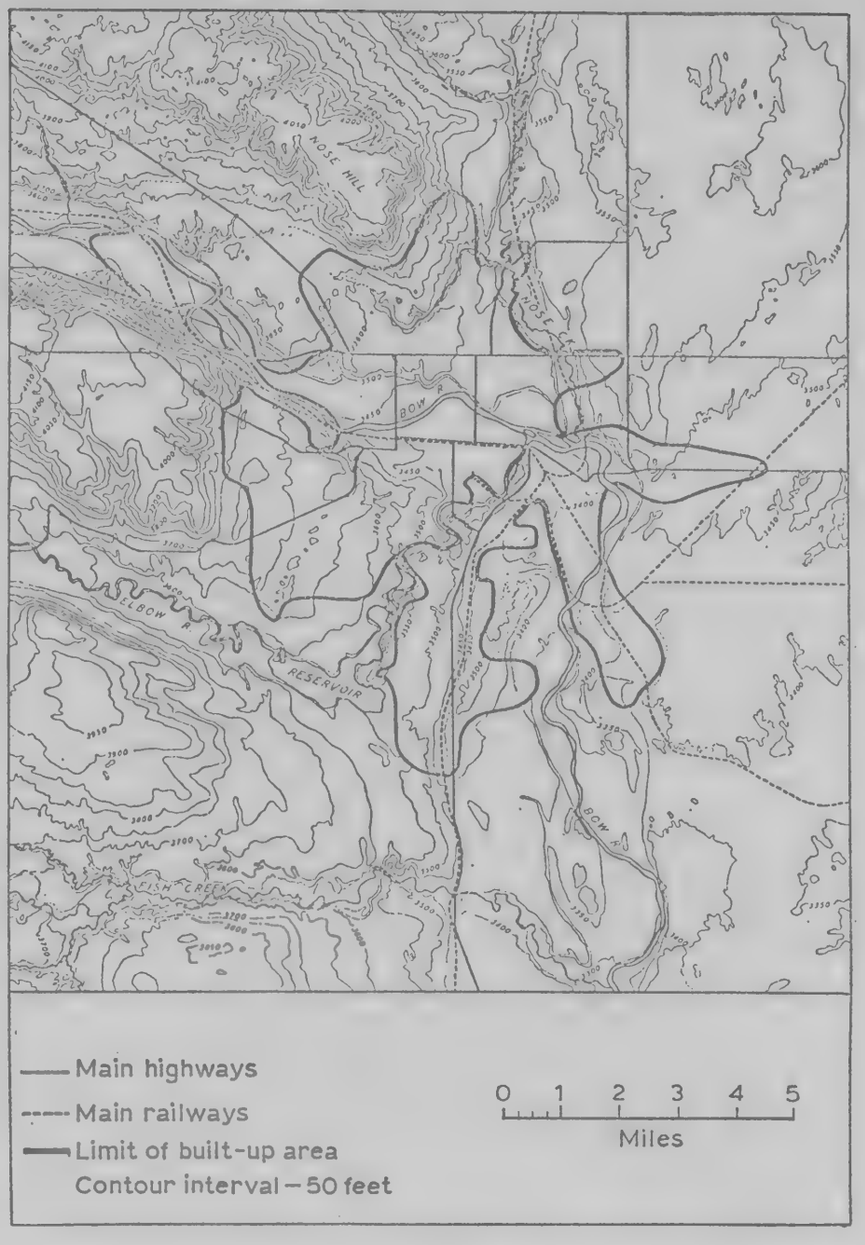


Fig. 21.7 The setting of Calgary and the relationship of the limits of the built-up area to terrain and transportation facilities

Hartmann then emphasised the need for a strong radial component if pure concentric patterns were to be developed in the city. His argument was not

demonstrated by analytical reference to any specific city, and so for such a study we can turn to the work of P. J. Smith which is reported in 'Calgary: A Study in Urban Pattern' (1962).[1] *Calgary, it should be noted, is situated at the foot of the Rocky Mountain cordillera in Alberta, Canada, and in eighty-five years has developed from a small police outpost to a city of 265,000 people serving a region of 75,000 square miles. It is remarkable in that three-fifths of its growth has occurred since 1948. As to any recognition in this city of the Burgess concentric zone theory, Smith has written:*

The two innermost of Burgess's five concentric zones, the central business–wholesaling–light manufacturing core and the surrounding residential zone in transition, are readily apparent in Calgary, but the remaining zones are by no means so clear. The outermost commuter zone does have a partial development, with 'exurban'[2] homes scattered along the major highways to the north-west and south of the city, but it would be impossible to differentiate zones of workingmen's homes and better-class residences as postulated by Burgess. These residential types exist but not in a double concentric pattern and, as Burgess recognised, the broad concentricity that does occur is disrupted by the radial cross-zones of industry. As far as Calgary is concerned, then, the widespread criticism of the theory as a oversimplification of urban form is valid.

Notes

1. *Economic Geography*, vol. 38, pp. 321, 323.
2. The term 'exurban' has been borrowed from Spectorsky (A. C. Spectorsky, *The Exurbanites* (New York: Lippincott, 1955) but as used in this paper it is not intended to have any sociological or psychological implications. Instead, it is used simply to describe those residences of commuters, usually of high income, which are scattered through rural areas in too loose a pattern for the term 'suburb' to be accurate.

If Smith implies criticism of this model, we can turn to the second type we have to consider, the sector model. This is developed on the assumption that the internal structure of the urban area is to a large extent determined by the disposition of routes radiating from the city centre. Variations in accessibility between radials causes marked sectoral variations in the land value surface and correspondingly an arrangement of land use sectors can be identified. We have already noted in passing Yeates's reference to this in his work on Chicago. Under the heading of 'Sectoral Variations' he wrote[1] *about:*

. . . the emerging importance of sectoral variations, a possibility that was suggested at the conclusion of the visual analysis of the land value maps (Figs 21.2 and 21.3). The broad areas of land

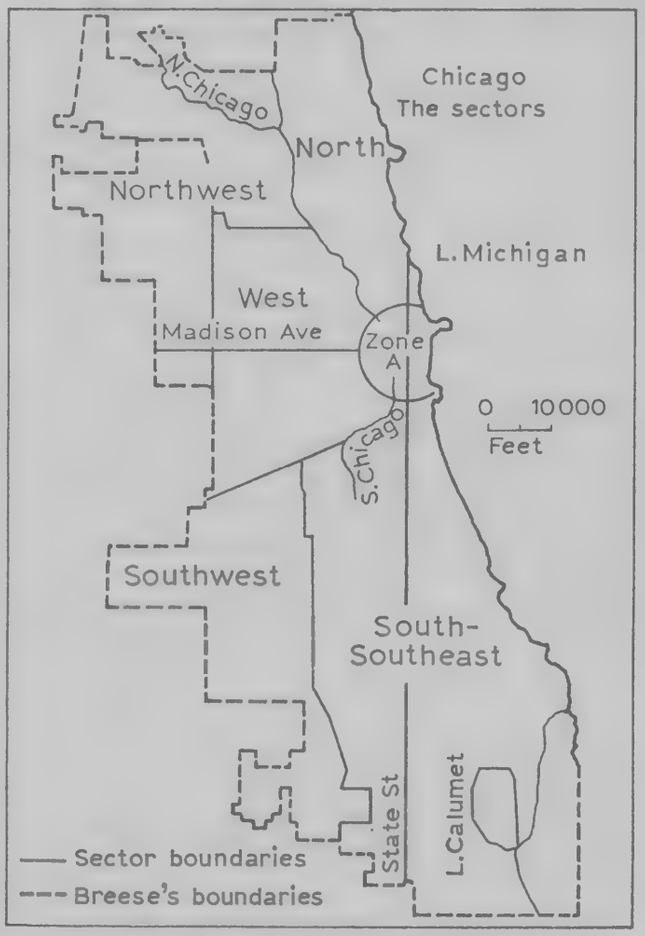


Fig. 21.8 Sectors in the analysis

values above and below the logarithmic mean suggest a number of sectors. Fig. 21.8 shows the location of these sectors, which are:

(i) *The central business district and vicinity (Zone A):* This zone includes all points within 1·5 miles of the intersection of State and Madison, thus conforming with Zone A. It is an area

largely dominated by commercial buildings, hence population densities are extremely low.

(ii) *The North sector:* This includes an area of high land values bordering Lake Michigan with a steep decline in the land value surface to the west. It is a sector of generally high income (median family income by census tract above $7000), and excellent central business district oriented transit facilities.

(iii) *The North-west sector:* The North-west sector also demarcates an area of values greater than the logarithmic mean, where land values related to housing have exhibited the most recent growth.

(iv) *The West sector:* The West sector has developed as an area of consistent, low land values characterised by the most recent and rapid extension of non-white neighbourhoods. This area is well served by transit and expressway facilities.

(v) *The South-west sector:* This represents an area of most recent growth in land values related to semi-suburban developments. The sector has always been remote from rapid-transit facilities.

(vi) *The South-south-east sector:* The South-south-east sector displays a number of important characteristics. Much of the northern part of the area has been well served by rapid transit since the turn of the century. Also, it has been the location of well-established non-white communities for four decades. Along the margin of Lake Michigan there exist pockets of high income (for example, around Jackson Park). Finally, in the far south-east around Lake Calumet, there is a broad area of heavy industrialisation. This whole sector embraces a large area of land values less than the logarithmic mean of the land value distribution in 1960, with pockets of higher values demarcating the high-income areas bordering Lake Michigan.

Note

1. M. H. Yeates, 'Some Factors Affecting the Spatial Distribution of Chicago Land Values, 1910–1960', *Economic Geography*, vol. 41 (1965) 66.

The sector model was, however, initially proposed by H. Hoyt[1] *when in 1939 he hypothesised that similar land uses concentrate along a particular radial route from the city centre to form sectors, as shown in the top right of Fig. 21.4 above. P. J. Smith,*[2] *in his consideration of urban land use in Calgary, examined Hoyt's model to see how far it explained the situation in that city:*

The sector theory is also, in some respects, an oversimplification, but at least the sector diagram of Calgary (Figs 21.9–21.11) includes most of the major elements of the land use pattern in a way that would not have been possible in a concentric zone diagram. And it recognises the essential dynamics of Calgary's

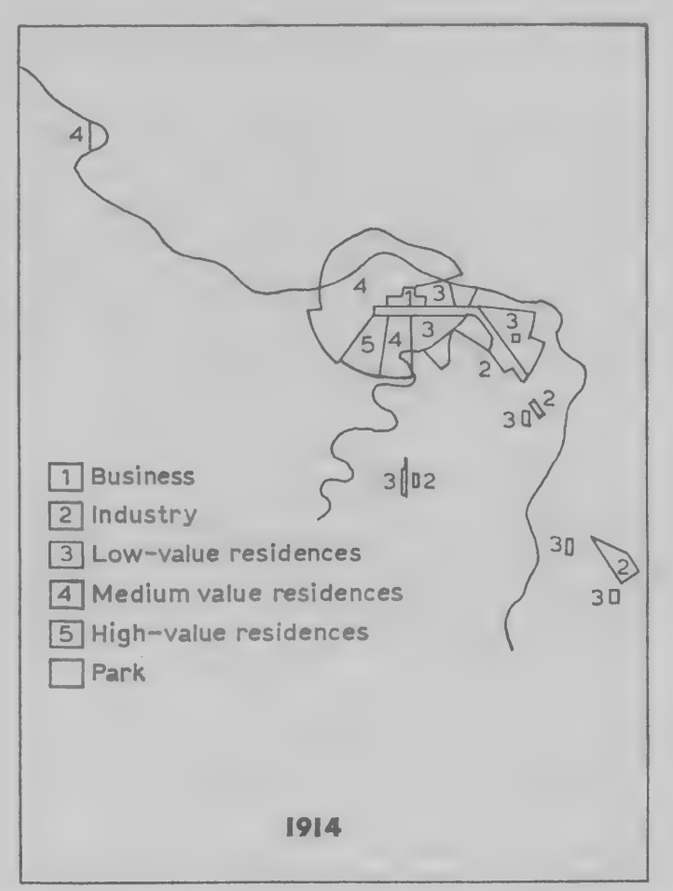


Fig. 21.9 Hoyt's sector theory applied to Calgary's land use pattern, 1914, when the city population was approximately 50,000.
Source: Assessment records, City of Calgary

growth, the outward progression of generalised land use types along radial lines of transportation.

Hoyt's studies, of course, were directed at residential areas and particularly at the identification of the growth patterns of high-rent and high-grade areas which he considered to have the dominant influence on the direction of residential growth. That is, the high-grade housing, through its social prestige and effect on land values, would draw the best of the medium-grade housing close to it while the low-rent and low-grade housing would be as far removed as physically possible. In Calgary, in general terms, this is true. The greater part of the city's high-cost housing is to be found in a narrow wedge exhibiting most of the characteristics claimed by Hoyt: a definite origin in the old, central housing district; growth along the narrow, very beautiful Elbow Valley towards open country where the homes of Calgary's wealthy 'exurbanites' are to be found; and a good radial thoroughfare as the axis of the sector. Conversely, much of the low-cost housing now appears in juxtaposition with industrial sectors, largely as a result of the depreciation and deterioration of dwellings consequent upon industrial expansion. The low-cost zones are by no means continuous, however, and new housing of better quality very often adjoins the industrial sectors. This reflects the fact that additions to the low-cost housing stock, apart from a limited scattering of rental row-housing units, are not being provided in the city, and that only in the three suburban towns, where land prices are still comparatively low, can adequate homes be purchased in the $10,000 to $12,000 range.

The peripheral tendency in the distribution of low-cost housing supports two of Rodwin's major criticisms of the sector theory:[3] first, that the empirical generalisations on which the sectors are based conceal the many great variations from the pattern, and second, that the theory overstresses nineteenth-century free market residential trends. For example, it makes no allowance for the circumstance that even the occupants of the cheapest forms of housing will have cars and will be able to travel beyond the fringes of the built-up area to obtain cheap land.[4] A comparable situation developed in the 1930s when peri-urban slums and near-slums emerged in the low-tax, unsupervised areas just outside the city boundaries but close to the termini of streetcar lines. Several of these fringe developments are now incorporated in the city and remain as incongruous pockets within the general sector of medium-grade housing. Similarly, pockets of very high-quality residences, impossible to identify on a sector basis, can also be recognised within this broad zone, usually in valley-side areas with scenic advantages.

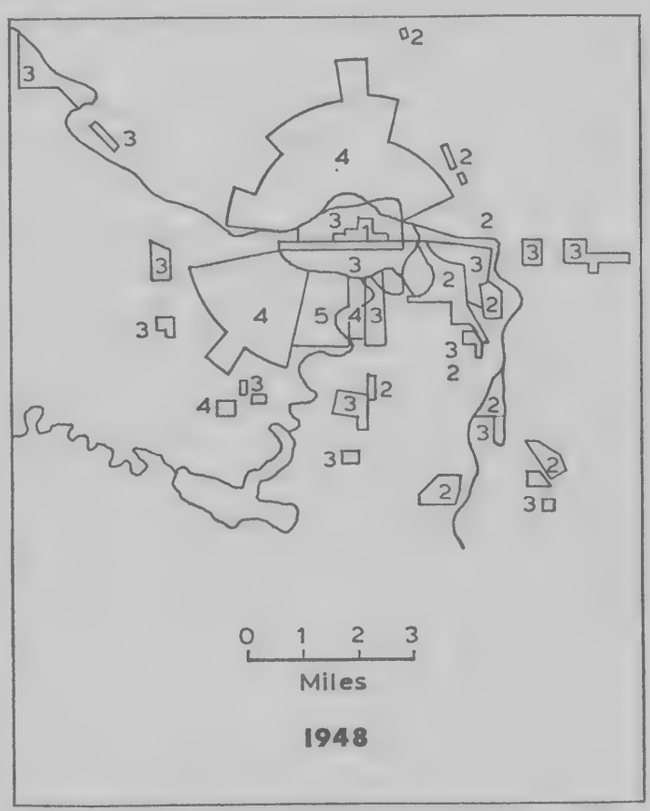


Fig. 21.10 Hoyt's sector theory applied to Calgary, 1948, when the metropolitan population was approximately 105,000.
Source: National Topographic Map Series

Rodwin also has criticised the emphasis on high-quality housing as setting the pattern for residential growth, another criticism valid for Calgary. How could it explain, for instance, the extensive recent expansion of medium to good-quality residences to the north-west and north, in an area isolated from the south-west quadrant of the city by the Bow River? Growth there is not related to the concentrated wedge of highest-quality housing but to existing highways,

to a planned programme of service extensions, to views of the mountains, and, most important of all, to the progressive development of a very large area of moderately priced, publicly owned land. Public interest in land development has, in this area, been more of a factor than free market forces in directing urban growth.

As a more general application of this criticism, it must be noted that the stress on high-quality housing conceals the significance of a more basic locational control, viz. industrial land use. Apart from some early small-scale housing developments in the Bow Valley, residential land use has avoided areas with potential industrial value, a circumstance now given legal force through Calgary's Zoning By-law.

These criticisms set definite limits on the validity of the hypothesis as a precise model of urban form and growth but they do not deny its basic worth. The schematic representation of the sector pattern greatly simplifies the process of understanding the city, particularly if Hoyt's original concept is enlarged to include the sectors of industries and highway-oriented businesses. The schematic base has the twin advantages of forming a useful descriptive tool through which the major features of the urban skeleton are exposed and of establishing a starting point for more detailed analysis. By first identifying the basic pattern and then accounting for the variations from the pattern, it is possible to obtain a satisfactorily complete interpretation of the city's internal form.

It will be apparent from the criticisms, however, that in the case of Calgary the sector theory is less adequate as an explanation of the forces of growth than as a description of present form. In general terms, the major sectors are being extended outwards, following transportation routes, but the reasons for the growth patterns of the various land use types are rather more complex than suggested by Hoyt and there are many land use elements which do not fit into the sector extension pattern. These elements have been taking shape for many years, since the peri-urban slums of the 1930s, and it is becoming increasingly evident that new trends in land use location are rapidly weakening the sector pattern that has been so characteristic of Calgary from the time the nascent settlement began to assume a definite form.

The disruptive elements have their origins in market forces and in large-scale public uses. Market forces have produced three principal features: the dormitory towns with their cheap land and low-cost housing; dispersed commercial facilities in compactly designed shopping centres or in strip zones along the provincial thoroughfares; and embryonic industrial nuclei to the south and north-east of the city. The major public intrusions are the airport,

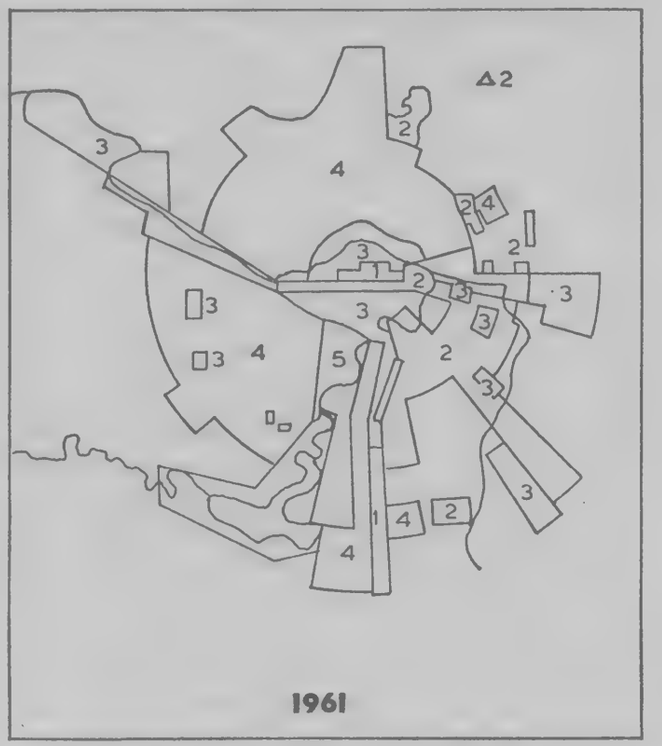


Fig. 21.11 Hoyt's sector theory applied to Calgary, 1961, when the metropolitan population was approximately 270,000.
Source: Field survey

which is both a barrier to private development and a possible nucleus for specialised commercial and industrial functions; the new university, which has stimulated proposals for high-quality housing; and the military establishments which have already greatly hindered the spread of residential development in the south-west part of the city.

There is strong evidence to suggest, then, that Calgary's sector pattern is now breaking down into the much more flexible multiple nuclei pattern, as elaborated by Harris and Ullman. The tradi-

tional nuclei, the central business and wholesaling district and the valley industrial zone, are still dominant in business location, but, in response to frictions of space, the special requirements of highway businesses, and the promotional campaigns of industrial estate developers, new business nuclei are assuming major importance. Undoubtedly, this trend will continue, particularly if the residential areas continue their low-density sprawl.

Notes

1. *The Structure and Growth of Residential Neighbourhoods in American Cities* (Washington: Federal Housing Administration, 1939).
2. 'Calgary: A Study in Urban Pattern', *Economic Geography*, vol. 38 (1962) 323–6, 328.
3. L. Rodwin, 'The Theory of Residential Growth and Structure', *Appraisal Journal*, Vol. 18 (1950) 295–317.
4. Hoyt recognises the occurrence of peripheral low-value housing but he also makes the claim that 'the shack fringe of the city is usually in the extension of a low-rent section' (Hoyt, op. cit., p. 121). In the Calgary area, this is justified only for the town of Forest Lawn which began as a shack town and has been improved only through the careful planning controls and building supervision of the late 1950s. The same applies to the towns of Bowness and Montgomery and the other fringe developments to west and south of the city, but these are all extensions of medium-rent sectors.

Smith mentioned the nuclei pattern of land use, which brings us to the last of the three models to be considered. This moves away from the concept of the development of urban land use patterns about a single centre. As C. D. Harris and E. L. Ullman suggested in their paper on 'The Nature of Cities' (1945):[1]

In many cities the land use pattern is built not around a single centre but around several discrete nuclei (see lower left-hand side of Fig. 21.4 above). In some cities these nuclei have existed from the very origins of the city; in others they have developed as the growth of the city stimulated migration and specialisation. An example of the first type is Metropolitan London, in which 'The City' and Westminster originated as separate points separated by open country, one as the centre of finance and commerce, the other as the centre of political life. An example of the second type is Chicago, in which heavy industry, at first localised along the Chicago river in the heart of the city, migrated to the Calumet District, where it acted as a nucleus for extensive new urban development.

The initial nucleus of the city may be the retail district in a central-place city, the port or rail facilities in a break-of-bulk city, or the factory, mine or beach in a specialised-function city.

The rise of separate nuclei and differentiated districts reflects a combination of the following four factors:

1. Certain activities require specialised facilities. The retail district, for example, is attached to the point of greatest intracity accessibility, the port district to suitable waterfront, manufacturing districts to large blocks of land and water or rail connection, and so on.
2. Certain like activities group together because they profit from cohesion. Retail districts benefit from grouping which increases the concentration of potential customers and makes possible comparison shopping. Financial and office-building districts depend upon facility of communication among offices within the district. The Merchandise Mart of Chicago is an example of wholesale clustering.
3. Certain unlike activities are detrimental to each other. The antagonism between factory development and high-class residential development is well known. The heavy concentrations of pedestrians, automobiles and streetcars in the retail district are antagonistic both to the railroad facilities and the street loading required in the wholesale district and to the rail facilities and space needed by large industrial districts, and vice versa.
4. Certain activities are unable to afford the high rents of the most desirable sites. This factor works in conjunction with the foregoing. Examples are bulk wholesaling and storage activities requiring much room, or low-class housing unable to afford the luxury of high land with a view.

The number of nuclei which result from historical development and the operation of localisation forces varies greatly from city to city. The larger the city, the more numerous and specialised are the nuclei. The following districts, however, have developed around nuclei in most large American cities.

The central business district. This district is at the focus of intracity transportation facilities by sidewalk, private car, bus, streetcar, subway.

The wholesale and light-manufacturing district. This district is

conveniently within the city but near the focus of extra-city transportation facilities.

The heavy industrial district. This is near the present or former outer edge of the city. Heavy industries require large tracts of space, often beyond any available in sections already subdivided into blocks and streets.

The residential district. In general, high-class districts are likely to be on well-drained, high land and away from nuisances such as noise, odours, smoke and railroad lines. Low-class districts are likely to arise near factories and railroad districts, wherever located in the city.

Suburb and satellite. Suburbs, either residential or industrial, are characteristic of most of the larger American cities. The rise of the automobile and the improvement of certain suburban commuter rail lines in a few of the largest cities have stimulated suburbanisation.

Note

1. *Annals of the American Academy of Political and Social Science*, vol. 242, pp. 14–16.

It should be remembered that the three models discussed so far, the concentric, the sector and the nuclei, are not mutually exclusive and elements of all three might be expected in cities. Smith recognised this in Calgary.

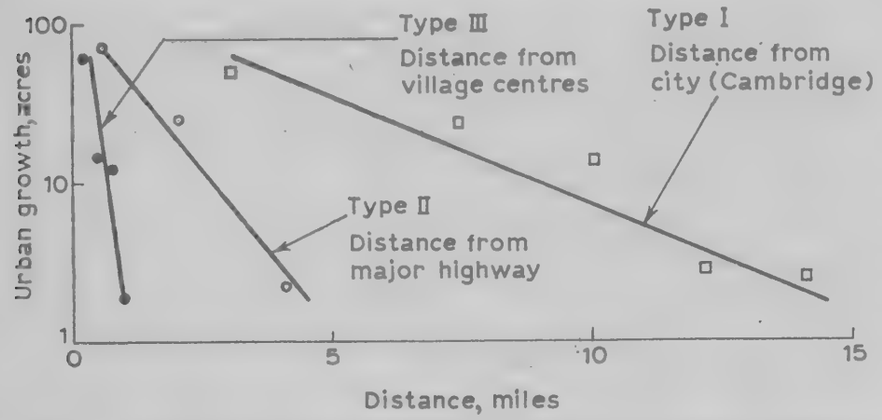


Fig. 21.12 Growth-distance relationships in urban conversion of land in southern Cambridgeshire.
Source: Field survey by D. E. Keeble and P. Haggett, 1964

More recently, P. Haggett in 'Location Analysis in Human Geography' (1965)[1] *has noted evidence from the south of Cambridge, where:*

the spread of housing since World War II shows three distinct gradients (Fig. 21.12): (i) a gentle gradient about the main city, in line with the concentric-zone model; (ii) a steeper gradient about the major highways, in line with the sector model; and (iii) a very steep gradient about the outlying villages, in line with the multiple-nuclei model.

Note

1. pp. 180–1.

All three models are useful in generalising about the internal structure of cities. However, in attempting to analyse the land use pattern of the urban area, one must not lose sight of the fact that land use has a dynamic quality. Patterns change under the influence of certain forces which accelerate or slow down the movement towards overall location efficiency. In the last analysis though, the key to the landscape pattern in an urban area remains transportation and its concomitant accessibility, and although these are not responsible for all changes in urban patterns, they occupy the central, universal role.

The notion of the urban land use dynamic and its underlying causal forces can be best understood in the context of Calgary from Smith's research.[1]

In the view of urban land economists, the emergent changes in the distribution and arrangement of private land uses in Calgary will be in response to a need for improvements in the efficiency of the urban pattern. To quote Ratcliff, under conditions of perfect competition 'each site becomes occupied by that activity which can use it most efficiently', with the result that, even under the normal condition of imperfect competition, 'there is an observable tendency for the natural economic forces to create an urban pattern which is relatively efficient in its basic space relationships'.[2] The efficiency of the land use pattern is not static, however, and the locational attributes of particular areas for particular types of land use will change in response to such factors as increases in the size and functions of the urban area; improvements in technology,

as evidenced in new techniques of transportation or new forms of industrial plant layout; and improvements and extensions to the public facilities needed in land development, notably thoroughfares, railways and utilities. The land use pattern, then, must be visualised as being in a constant state of flux, for ever adjusting to changing circumstances, working towards but not necessarily attaining maximum locational efficiency. There are, of course, many reasons for this failure. The inevitable delay between an external change and its effect on land use, the inertia of an established land use pattern, the failure or inability of a local government to provide necessary public facilities are but some of the many important factors which may retard the evolution towards locational efficiency, even to the point of creating inefficiency in the urban pattern.

In applying this concept to Calgary, it is necessary to examine those inefficiencies in the land use pattern which are leading to the breakdown of the sector form. There would appear to be four principal ones.

As a result of the rapid spread of low-density residential areas in the past decade the dimensions of the city, the frictions of space, have become too great for the central business district to act as the sole retail and service centre. In 1950, the city's population was 120,000 and the radius of the built-up area averaged only 2·5 miles. Only six commercial nuclei, all on major traffic routes, were sufficiently well developed to be regarded as shopping centres, and their facilities were essentially of the local grocery–drug store–service depot variety (Fig. 21.13). Even as late as 1956, there was no very great change to this pattern. Commercial strips were being extended and supermarkets were making their appearance, but land set aside for shopping centres in the new outlying residential neighbourhoods was still vacant. By 1961, however, the metropolitan population had increased to 265,000, the urban radius to an average of five miles and the number of dispersed shopping areas to sixty, including two regional centres each with a department store and approximately fifty smaller stores, service facilities and offices. It would almost seem that the 1956 population of 180,000 represented a critical limit for the efficiency of the central business district, by no means marking the beginning of a decline in central retail functions – quite the contrary – but clearly differentiating the specialised retail character of the central area from the localised centres efficiently dispersed for day-to-day convenience of consumers.

Both the central wholesale district and the valley industrial district are oriented to railway transportation and both have become inefficient with the modern emphasis on truck transportation. The wholesale district has become inefficient for three reasons:

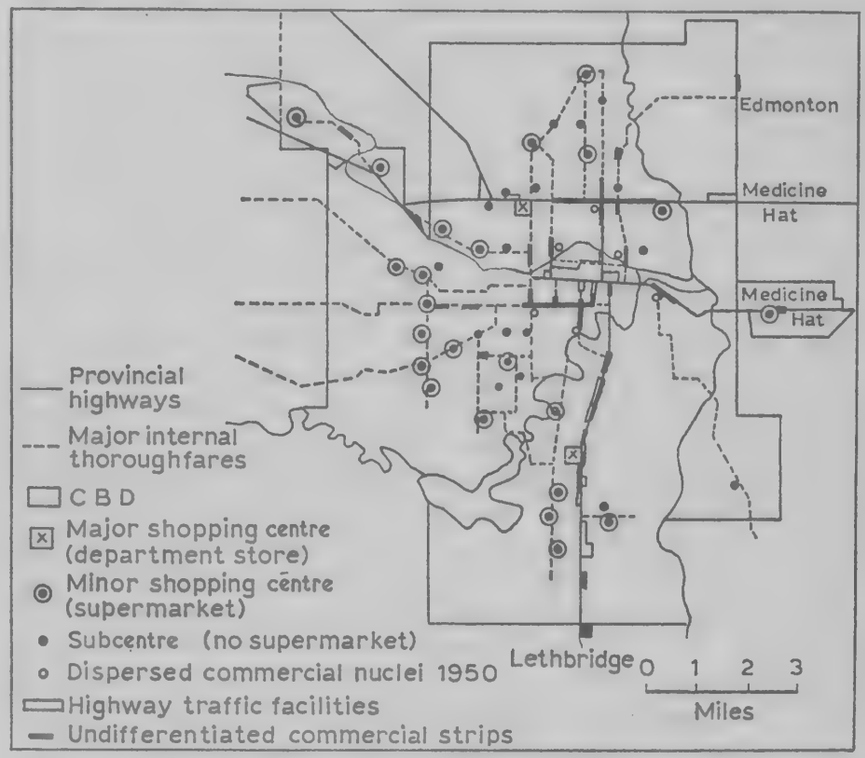


Fig. 21.13 The distribution of commercial development in the Calgary metropolitan area, January 1961

the old multi-storey stores and warehouses were not designed for modern techniques of handling goods; the narrow streets and the lack of loading spaces off the streets greatly handicap the manœuvring of trucks and the transfer of goods; and the dispersal of commercial facilities means that a central location for the distribution of wholesale wares is no longer necessary. The industrial valley has become inefficient principally because its street pattern, a standard residential grid badly disrupted by railway lines and

topographic features, is hopelessly unorganised and a serious hindrance to efficient traffic movement.

The first move away from these areas came in 1955 when two planned industrial estates were opened on municipally owned land between the Bow Valley and the main highway to the south. These areas were carefully designed and fully serviced and provided excellent access to highway and rail facilities and to all parts of the city, locational advantages which were promptly recognised by both new and relocating businesses. The municipal estates account for the bifurcated form of the industrial sector and their success has stimulated the private interest in industrial land development that accounts for the emergence of the new nuclei mentioned earlier. It is most unlikely that the industrial sector will ever be displaced in importance, particularly since a major scheme of thoroughfare construction is now in progress, but the new nuclei have advantages of access which will assure their remaining competitive on the basis of locational efficiency.

Two more specialised types of businesses which have undergone a locational shift as a result of the rapid increase in highway transportation are the provision of accommodation and ancillary services for through-travellers and the sale of new and used automobiles. Before 1950, both groups of activities were confined almost exclusively to the central area, but between 1950 and 1961 only one hotel, one motel and a small number of used-car lots were added. During the same period, and more especially since 1956, four hotels were constructed in outlying areas, motel concentrations appeared along all the provincial highways and both new and used car agencies became localised in a two-mile-long strip flanking the city's main southern access.

The increased land prices and municipal taxes, which have resulted from the boom in the central business district, have contributed largely to the locational disadvantage of that area, particularly for motel operators and used-car dealers, but this is by no means the only explanation for the new pattern. It is generally believed that the needs of highway traffic can be satisfied most efficiently along the highways approaching the city, as long as the motels are not too remote from the city's main attractions; certainly competition between the businesses catering to highway travellers can be conducted most efficiently on the access routes. The car dealers, of course, are not interested in transient traffic but their highly competitive business does seem to benefit from the continuous strip form of development that enables both buyers and sellers to make ready comparisons of prices.

The final major inefficiency, the extreme shortage, within the City of Calgary, of land available for low-cost housing, has already been discussed. This inefficiency, combined with the ready availability of cheap automobiles, has resulted in the rapid growth of the three suburban towns, forming, in effect, three nuclei of low-cost residential divelopment. The combined populations of these towns increased from 5500 in 1951 to 25,000 in 1961.

These, then, are the principal locational changes of private land uses apparent to date. They represent the beginning of a major breakdown of the sector pattern as a result of the increasing dimensions and the expanding functions of a very dynamic urban area. They reflect the way in which urban growth can produce inefficiencies in an established pattern, principally through its effect on ease of transportation and on land costs, or site rentals as the land economists prefer to say. They reflect the way in which technological changes, new forms of transportation and new techniques of mass handling of industrial goods can make existing land use areas functionally obsolete, unless large-scale internal improvements are made to buildings and thoroughfares.

The changes to the land use pattern are, as yet, still in a nascent stage. The sector pattern still describes Calgary's form as completely as could be expected of any idealised model. In fact, some of the land use changes, the municipal industrial estates and the ribbon commercial development, can be regarded as elaborations and extensions of the sector pattern. The other changes, however – the private industrial estates, the dispersed nuclei of retail and highway commercial facilities and the peripheral low-cost housing – illustrate definite recent departures from the sector pattern. When these are combined with the large-scale public uses mentioned earlier, it becomes apparent that Calgary's arrangement of land uses has assumed at least the beginnings of a pattern focusing on several nuclei.

For future years, it seems reasonable to anticipate that this trend will be intensified. In one recent study, it has been postulated that Calgary's metropolitan population will be greater than 600,000 by 1980, an increase which will likely cause the urban area to be expanded two and a half times and will certainly produce even

greater dispersal of business functions.[3] New inefficiencies will undoubtedly appear, locational advantages will change and more and more urban functions will benefit from nucleated dispersal. For example, until recently no major office building could be constructed outside the central business district, because of zoning restrictions, but as a result of increased business feeling that some types of offices could operate more efficiently in other areas, largely because of better access and more favourable land costs, the Zoning By-law was amended. One large office block was constructed out of the central area in 1960 and proposals for others have been made. This is just one evidence of future demands for dispersal.

In conclusion, it should be stressed that transportation is as much the key to the land use pattern as it ever was. The weakening of the sector form has not changed this. Transportation is by no means the only determinant of land use or the only factor accounting for locational advantages or disadvantages: site rentals, financial and physical limitations to utility extensions, public actions in land use and land-use controls, the activities of real-estate promoters, competition between businessmen and the social needs and desires of home-owners all have a very important share in deciding the arrangement of land uses within any city. But none of these occupies the central universal role of transportation. Transportation may not be responsible for every change in the urban pattern but, on the other hand, no change will be possible unless reasonably efficient and convenient transportation can be provided.

Notes

1. *Economic Geography*, vol. 38 (1962) 326–9.
2. R. U. Ratcliff, 'The Dynamics of Efficiency in the Locational Model', in *The Metropolis in Modern Life*, ed. R. M. Fisher (New York, 1955).
3. P. J. Smith, *A Study of Calgary's Past and Probable Future Growth* (Calgary, 1959).

THE PROBLEM OF SETTLEMENT PATTERN ANALYSIS

The pattern made by the way in which man is distributed across the surface of the land masses of the world is of first importance to the geographer and is central to his contribution to any analysis of man and his behaviour. So far in this group of chapters, some attempt has been made through the use of basic models to contribute to an understanding in spatial terms of his use of rural and urban land, but as yet no attempt has been made to account for the spatial relationship between these two uses – no attempt, that is, to account for the way in which urban settlements of varying types and sizes are distributed through the rural landscape. In an analysis of this problem the immediate difficulty which arises is one of definition. In former times when towns and cities were ringed by walls, the distinction in terms of settlement between what was rural and what was urban would have been clear. But today, although there is a tendency to think in terms of population as either rural or urban, simply because local government units are at present divided on this basis, the use of such a classification can be misleading. Would it be right, for example, to consider a large village as rural in character if an attempt was being made to analyse the relationship between the distribution and size of settlements in an essentially agricultural landscape? Certainly it would not. Another possibility in attempting to distinguish between rural and urban populations might lie in the choice of a certain population density per square mile as a dividing line. The insoluble difficulty, however, would be to agree a figure which would suit a wide range of countries and environments.

Therefore the only realistic way in which to analyse the relationship between urban settlements of varying sizes and the rest of the rural landscape is to abandon attempts at a population definition or indeed any other definition and think in terms of what is knows as 'central place theory'.

CENTRAL-PLACE THEORY

This seeks to relate 'central places' to their regions or hinterlands; it defines a central place as a settlement that provides one or more services for the people living outside it. The great advantage of this theory, which establishes a relationship between settlement nodes and the areas they serve, is that it can be universally applied. It provides a framework by which settlement systems can be studied throughout the world. However, the central-place concept, like the other models we have used in this group, is not a perfect one since in its explanation of the size and spacing of settlements it accounts only for the intensity and spacing of service activities which are centrally based and provides an understanding of the way in which urban units have evolved and are changing only as service centres. It has to be remembered that towns and other urban centres have within them other activities which cause them to grow and evolve besides that of making a service provision for the surrounding area. A moment's thought about the two earlier chapters in this group must make this apparent. There are also other problems with regard to this concept which will be dealt with later. Certainly, whatever the imperfections, central-place theory has a validity which now justifies a more detailed consideration of the background from which it stems. For this purpose it is worth while to look at the recent and lucid writings on this subject by Peter Ambrose in his book 'Analytical Human Geography' (1969).[1] In examining the services provided by a central place, he pointed out that these can range from a post-box in a small hamlet up to an eminent surgeon working in a capital city whose 'region' might be considered as world-wide if his skills were such that he could draw patients seeking his services from any or all of the continents. Ambrose continued:

Between these two extremes exist a whole range, or *hierarchy*, of functions. In ascending order of importance this hierarchy would include (to take seven examples from many hundreds): a small general store, a public house, a ladies' hairdresser, a hardware shop, a secondary school, a department store and an opera house. This list is arranged in a progression from 'low-order' to 'high-order' goods or services and the significance of this is that normally (but not always) wherever a 'high-order' function occurs in a central place, then one or more of each of the 'lower-order' functions also occurs in the same place. Thus any centre (whether town, village or even suburban shopping parade) that has a hardware shop is also likely to have all the lower-order goods listed before it. Similarly, any large regional centre is likely to have all seven functions listed.

Another important feature of the list should be noticed. The population required to support each function tends to increase as one goes higher up the hierarchy. Whereas a small general store may exist with a 'market' of, say, two hundred regular users, a

hardware shop may require perhaps one thousand. A secondary school may require five thousand regular 'users' (assuming that it is built for five hundred pupils between eleven and eighteen and that people between these ages constitute 10 per cent of the local population). The number of people required to support a function is called the *threshold population*, and if this is not maintained the function may cease to operate. This can occur either when the local population decreases (which may cause a village shop to close in an area subject to rural depopulation) or when changes in taste decree that a smaller percentage of the population use the function (which partly explains the recent closure of many cinemas). Conversely, as local population, or more accurately local user population, increases above the threshold the function will tend to make larger profits. This may continue until a population approaching twice the threshold is reached, by which time someone will have realised that he may well be able to profit by offering the same function. Competition, or even possibly co-operation, then occurs between the two purveyors of the function.

Threshold populations for the different functions vary because our needs conform to a pattern. Some things, such as food, newspapers, petrol and cigarettes, are needed very frequently. Others, such as furniture, are needed less frequently, while really high-order functions, such as an opera house, may be visited only once a year. In the latter case, to be sure of filling an opera house seating one thousand people every week-night a threshold population of about 300,000 is needed. If, following a wave of enthusiasm for opera, people tend to average six visits to the opera per year, a threshold population of 50,000 will be sufficient to support the opera house. The expected frequency of use can therefore be seen to be an important aspect of threshold population, and some functions, for example funeral directors, may need an enormous threshold population, because individuals use them very infrequently. Others, such as tobacconists (and schools), can exist on a smaller threshold population, because a sizeable and predictable number of people use them daily.

Having considered these fundamental, but not too difficult, ideas we can return to the problem with which we started. How can we build up a theory, or a set of explanatory statements, which will account for the way in which central places of widely differing sizes have come to be distributed as they are in the landscape? At this point our thinking, while still being based on common sense and simple ideas about our needs, becomes spatial and involves some simple geometrical ideas.

We can begin by assuming in some mythical region a population entirely engaged in agriculture. Each family is completely self-sufficient and therefore indulges in no trade. For the sake of simplicity we can assume that the land is of a fertility that does not vary spatially and that the culture is so egalitarian that each family farms a plot of standard size. Thus the population density does not vary spatially either. One of the farmers, having an inventive turn of mind, discovers how to make an implement that will increase crop yields. After a while he realises that he can stop farming and make a living by exchanging this implement for food produced by others. He becomes the first central place in the region because people come to him with food and trade it for an implement. He begins to employ people and other buildings are added around his farm to house his employees. So far the central place has an entirely industrial function, but as more and more people live there, all dependent on exchanging a manufactured good for food from the farming areas, other functions are bound to arise. There must be book-keepers to keep the records and storekeepers to look after the stocks of machinery awaiting exchange. At a later stage other implements will be invented and the economy will have to devise some money token representing value, because trading in actual commodities rapidly becomes very unwieldy as the number of commodities increases above the original two (the implement and food). Banking, retailing, credit and insurance all evolve in due course. As the regional culture grows in sophistication, political and educational institutions, repositories for works of art and places for recreation will all become necessary and will be added. From an early stage in this process of accretion of new services, people will come from outside the manufacturing centre to make use of the functions it provides, so that as well as being an industrial centre it will have become a service centre. In this simple outline of events most of the essential elements of urbanisation are present. There could be a great deal of variation in the detail. For example, the first centre could have been oriented to trade between producers of different crops rather than to manufacturing. But the key ideas of functional interdependence between central place and surrounding region, an increase in the variety of central function offered and an

increasing percentage of the population not making their living from farming are all evident, as they have been historically in urbanisation.

The geographer has a very specific interest in this process. He is concerned to examine the principles that seem to govern the *location* of each new function as it is invented or emerges. There are many principles involved, but we will think only of three at this stage. One relates to the idea of *threshold population*. As each function is added, a certain number of people are required to support it. This normally means that they must be able to reach it from where they live. Movement always costs the individual something. It will therefore be easier to attract a threshold population to a function, whether it is a theatre, a shop or a museum, if the function is located in an existing population centre. Here people are already living at a higher density that they are in the farming areas. Therefore the aggregate distance that has to be travelled to bring, say, one hundred people to a shop every day is less if it is located in some existing centre than it would be if the shop were opened in a rural area.

The second principle also leads to the choice of existing central places as the location for new functions. Some functions are closely interdependent. The shops need to be near the factories producing the goods they sell. The bankers need to be near the manufacturers so that they can discuss business affairs more easily. The government needs to be near the bankers and the manufacturers so that they can co-ordinate activities more effectively. In fact there are savings to be made by all these institutions if they locate near the other institutions with which close contact is necessary. These savings are called *external economies* because they stem from factors outside the institution concerned and, in the cases mentioned, they are the result simply of building in one place rather than another; in making wise use of the existing spatial pattern of people and institutions.

These two principles alone do not give any indication about which particular central places, in a given region, will tend to attract new functions or how the pattern made by centres providing a particular function will evolve. To illustrate this we need to consider a third principle, that of *spatial competition*. Let us imagine a region in which central places of a more or less equal size are already established and are regularly spaced at a distance of about twenty miles from each other. In this region the people have traditionally solved the problem of storing their money by the rather crude expedient of stuffing it up the chimney in a sock. One day someone develops the idea of opening a bank where people can deposit money which is surplus to their immediate requirements. This spare money can, of course, be put to profitable use by the

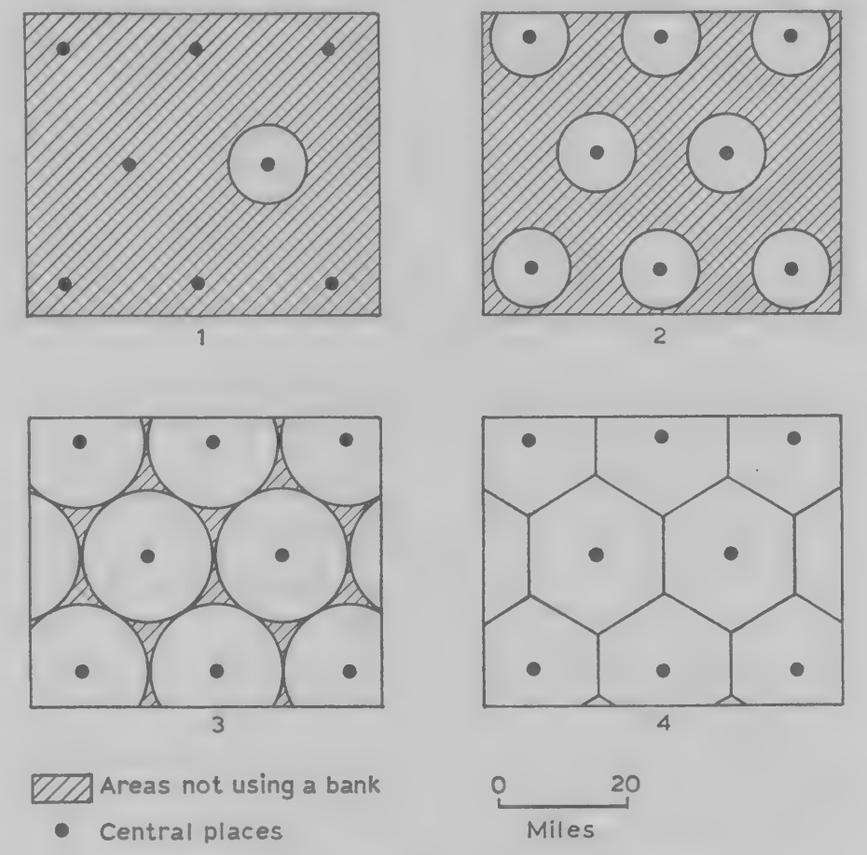


Fig. 22.1–4 Colonisation of a region by bankers

banker. The people are now faced with the choice of adopting the new method of storage, thus presumably gaining in safety and making some money in interest, or of keeping to the old method and avoiding the cost and inconvenience of the periodic journeys to the bank (which, for reasons already outlined, will be located in one of the central places). Let us assume that above a certain

journey distance, say five miles, the cost of journeys to the bank outweighs the advantages of the banking method. The bank can therefore attract as customers only those living in the central place plus those in the neighbouring countryside up to a radius of five miles from the bank (see Fig. 22.1).

As the banker is seen to prosper and grow rich, other people set up in the same business. They will each choose a central place and, to avoid unnecessary competition from other banks, each will choose a different central place until all central places have a bank (Fig. 22.2). Now, as the development of the transport network reduces the cost of movement and as the banks acquire greater expertise and reduce costs by using computers (this is, after all, a mythical region), consumers become willing to travel more than five miles to use a bank. As soon as it becomes worth travelling ten miles, the situation shown in Fig. 22.3 is reached. Very soon the people living in the shaded areas will want to use a bank, and if they all go to the nearest one each bank will be serving a hexagonal area as shown in Fig. 22.4. Now the whole region has been effectively 'colonised' by bankers.[2]

This situation can continue to develop in a number of ways. For example, it is likely that smaller central places, if they have not done so already, will begin to grow up at those points where three service areas meet. These are the points from which the cost of reaching existing central-place services is at a maximum. A new banker coming into the region may help this process by locating at one of these smaller centres. Whether he does so or not will depend upon the threshold population required for banking and upon the size differential between the smaller and the larger central places. Another development might be that one bank, which we could call bank X, finds a way of reducing costs substantially (perhaps by selling its computer) so that it becomes worth while to make a longer journey to reach it. People just over the boundary of this bank's service area are thus tempted to transfer their business from their existing bank to bank X. This has the effect of distorting the perfect hexagonal scheme of service areas and the more efficient bank enlarges its own service area at the expense of the banks around it.

Notes

1. pp. 121–6.

2. J. A. Everson and B. P. FitzGerald, in *Settlement Patterns* (Longmans, 1969) pp. 103–4, offer the following note on the hexagonal form of service area:

> The hexagonal form of these service areas is mathematically the most efficient. This subject is often referred to as elementary packing theory. One can judge the efficiency of packed shapes in two ways. The first concerns the efficiency of movement, i.e. as measured by the distance from the centre to the most outlying parts of its boundary; and secondly by the efficiency of the shape's bounding limits, i.e. the measure of the length of the perimeter for a given area of territory.

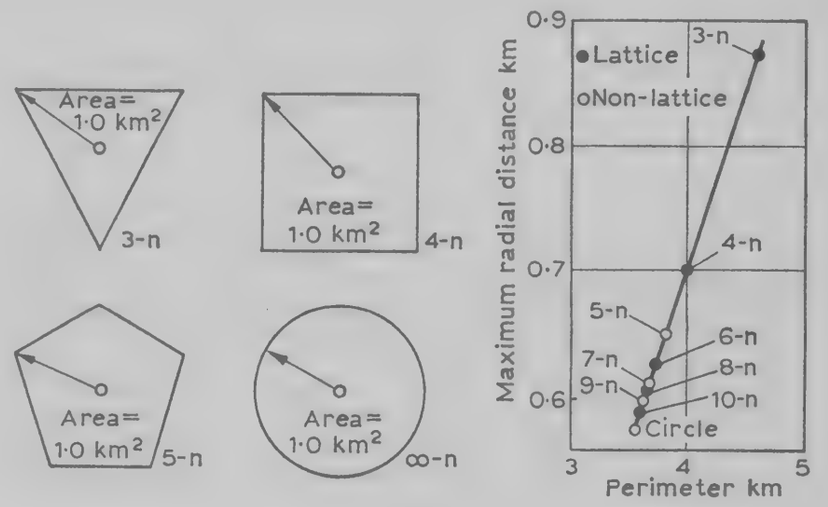


Fig. 22.5 Efficiency of alternatives types of regular polygons in relation to distance from centres and perimeter lengths (after Haggett, *Location Analysis in Human Geography*, 1965)

With a given unit of area, the circle is the most efficient from both these points of view. However, a circle is a shape that will not pack without leaving spaces between neighbouring circles. The only regular polygons that will allow packing (i.e. will allow the formation of regular lattices) are the triangle, the square and the hexagon. Given that the area of each of an equilateral triangle, a square, a regular hexagon and a circle is one square kilometre, Fig. 22.5 shows the maximum radial distance in kilometres plotted against the length of perimeter in kilometres. From this it can be seen that of all the regular 'tessellations', the hexagon retains most of the advantages of the circle, being about four-fifths as efficient as the circle.

The above quotation from 'Analytical Human Geography' provides the background to an understanding of central-place theory, a concept first developed by Walter Christaller in 1933 in his work 'Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen'.[1] *Christaller stated his theory to explain the 'size, number and distribution of towns' in the belief that 'there is some ordering principle governing their distribution'. It is interesting to note that he considered his theories 'could be designated as the theory of location of urban trades and institutions', to be placed besides von Thünen's theory of the location of agricultural production and the theory of location of industries developed by Weber.*

A direct abstract of the theoretical parts of Christaller's work was made by B. J. L. Berry and A. Pred for their book 'Central Place Studies':[2]

The crystallisation of mass about a nucleus is part of the elementary order of things. Centralistic principles are similarly basic to human community life. In this sense the town is a centre of a regional community and the mediator of that community's commerce; it functions, then, as the central place of the community.

Central places vary in importance. Those of higher order dominate larger regions than those of lesser order, exercise more central functions, and therefore have greater centrality. For all, however, 'the sum of the distances which rural residents travel to the central place is the smallest conceivable sum'.

The goods and services provided by the central place because it is central are known as central goods and services. Higher-order goods are offered at central places of higher order, and lower-order goods at places of both higher and lower order.

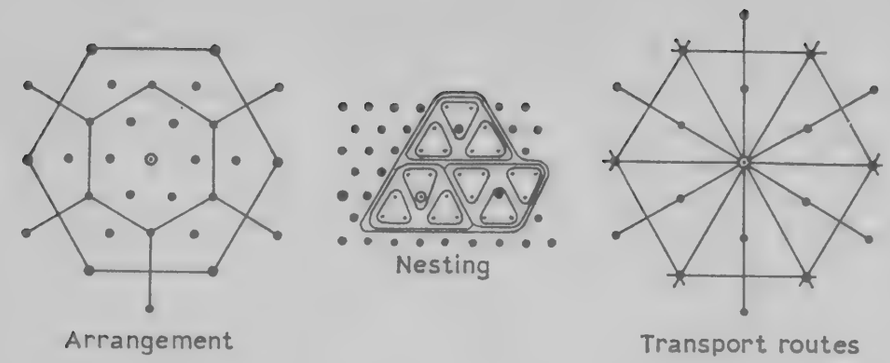
The region for which a central place is the centre should be called the complementary region; it includes relationships in both directions – town to country and country to town. Complementary regions are likewise of higher and lower order. They are hard to determine because they differ for different types of goods, undergo periodic and seasonal variation and consistently overlap neighbouring complementary regions at the periphery. Distance plays a vital role in any determination of complementary regions, especially 'economic distance', measured in time and cost. The range of a good is the farthest distance a dispersed population is willing to go in order to buy a good offered at a place. This range will take on a lower limit if there is competition from another centre.

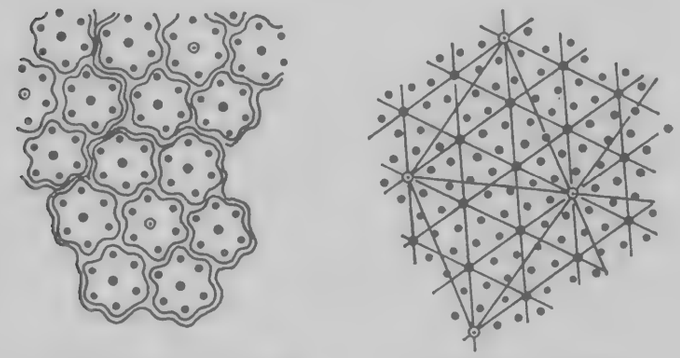
A decisive fact in the development of central places is the net income which inhabitants earn. There is a functional relationship between the size of a central place on the one hand and the sizes of the complementary region, its population and income on the other.

Central goods offered at a larger place have a larger range than those offered at smaller, and the fact that a central place is larger or smaller has an immediate influence on the range of a central good because more types of central goods are offered at a centre of higher order. The possibility that on a single trip several types of goods may be offered simultaneously has the same effect as a general price decline in goods offered by larger towns.

Every type of good has its special range, which differs at different central places and is not the same in all directions from the same centre, but varies according to objective and subjective economic distance. More basically, range is determined by (1) the size and importance of the centre and the spatial distribution of population, (2) the price-willingness of purchasers, (3) subjective economic distance and (4) quantity and price of the good at the central place. The range is actually a ring with an upper limit beyond which a good can no longer be obtained from a centre, and a lower limit which is determined by the minimum amount of consumption which is necessary before production or offering the central good will pay.

There is a system of central places comprising several size-types, determined in general by the spatial effects of the upper and lower limits to the range of central goods. According to this *marketing or supply principle*, if the assumption of a homogeneous plan with equal access in all directions is made, complementary regions become hexagonal, and lower-order centres and their complementary regions 'nest' within those of larger centres according to a rule of threes. In this system the relations between size, spacing, functions and hierarchical interdependence in the system of central places are determined precisely. Seven levels to the hierarchy are postulated from the level of hamlet to world-city, based upon South German evidence. This strict mathematical scheme is, of course, as imperfect as the simplifying assumption, and reality may be approached by recognising price differences, differences in





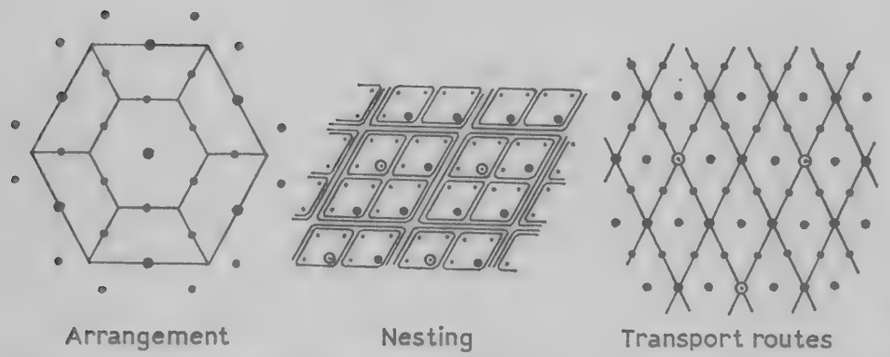


Fig. 22.6 The system of central places after (*top*) the marketing principle, (*centre*) the administrative principle, (*bottom*) the transportation principle

population distribution, etc. Note the strong parallel between size and frequency of central places and intensity of traffic, but it is the former which determine the latter, and not vice versa.

The system of central places developed on the basis of range of central goods used the assumption that all areas were able to be served from a minimum of central places; therefore the principle on which the system was developed can be called the *marketing or supply principle*. But there are other factors. *The principles of traffic* say that the distribution of central places is at an optimum where as many important places as possible lie on one traffic route between larger towns, the route being established as cheaply as possible. Complementary regions then 'nest' according to a rule of fours. Principles of traffic are fundamentally linear, those of marketing spatial. *The political-social (administrative) principles* are based upon ideas of separation of complementary regions for purposes of protection, or of distinction which implies clear-cut administrative control. In this case 'nesting' follows a rule of sevens. Fig. 22.6 depicts the patterns that result from the operation of each of the three principles.

The three principles determine, each according to its own laws, the system of central places. Two of them are economic and one is political. Under certain circumstances one or the other principle may dominate, but mostly they have to fight for predominance, a predominance which depends upon the operation of dynamic processes – of short-run, periodic and seasonal fluctuations, and longer-run secular changes.

Notes

1. Translated by C. W. Baskin, *A Critique and Translation of W. Christaller's Die zentralen Orte in Süddeutschland* (University Microfilms, Ann Arbor, Mich., 1957).
2. Regional Science Research Institute, Bibliography Series I, pp. 15–17.

It is important to understand from the above summary that the notion of variations in size of central places and variations in importance of the functions provided by each must give rise to a hierarchical system of hexagonal service centres with the service areas of successively larger centres imposed on the service areas of the smaller centres. As Christaller stated, this superimposition could occur in one of several ways depending upon which of

the locating principles (marketing, traffic or socio-political) was operating to greatest effect in the area. This argument has been spelt out in greater detail by J. A. Everson and B. P. F. FitzGerald in 'Settlement Patterns' (1969).[1] *But they begin by taking central-place theory back to basic principles in a way similar to that of Ambrose. Using a hypothetical landscape, they first assume a newly settled population to be mainly agricultural.*

We also assume that families tend to settle in groups, if only for the sake of companionship and defence against possible marauders. Thus small but largely self-sufficient hamlets are established, each with its group of families farming the land around it.

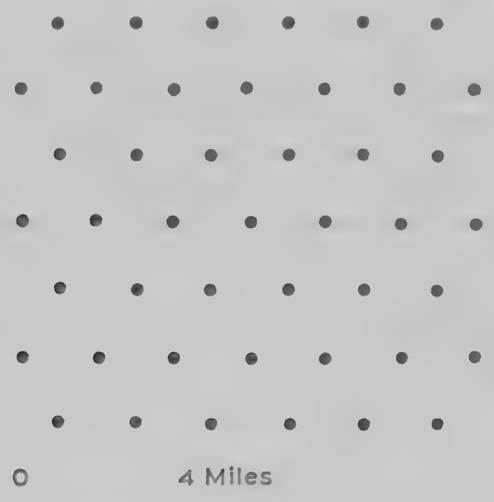


Fig. 22.7 Hamlets spaced two miles apart on an equilateral triangular grid

If it is assumed that the physical and farming characteristics of the area are everywhere the same, and that each settlement consisted of an equal number of farmers each gaining a sufficient livelihood from the land, then the distribution of hamlets seen in Fig. 22.7 will be expected. Here the pattern is perfectly regular, with each hamlet located at the apex of an equilateral triangle. In this case, the hamlets have been spaced at a distance of two miles from each other.

So far the hamlets have been assumed to be more or less self-sufficient. However, as advances are made in technology through time, it is safe to predict that some specialisation of production will occur, and that given reasonable roads some trading will develop between centres. Therefore it would be expected that the occasional hamlet would contain an enterprising group of people who would see the advantages of organising the exchange, or buying and selling of local produce. By doing this the group concerned would make a little more profit than it would do otherwise. Now if it is assumed that a two-mile journey either way is the farthest that a farmer would be willing to travel on foot, it can be seen that the hamlet with the enterprising traders will share in the trade of six of the surrounding hamlets, and of course its own. The trading area (worked on the basis of a day's travel) of this central place hamlet is seen to be hexagonal in shape. Other centres are seen to develop in a similar manner, as profit is seen to be gained in acting as a trading centre. If the trade of the equivalent of three hamlets is required to make a reasonable profit for the trader, the form that the hexagons must take would be that shown in Fig. 22.8. Here the hamlets around the central-place hamlet each share their trade between three trading areas. Thus there are six one-third

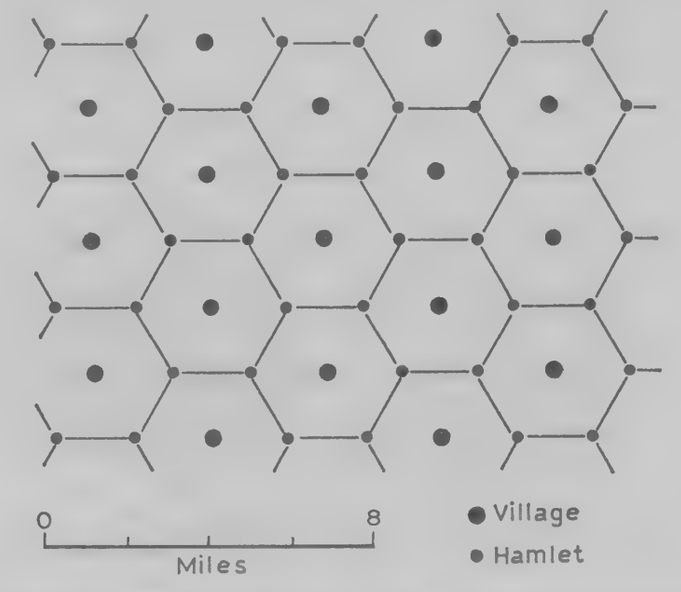


Fig. 22.8 Pattern of hamlets and villages service area governed by a $k = 3$ hierarchy

allegiances, and this added to the original hamlet makes a total custom of three.

In the hypothetical landscape with its pattern of hamlets and central-place hamlets, or villages, it can be seen that as more specialisation is achieved and as communication improves further, so the settlements will become less and less self-supporting. Several of the villages will specialise still further, introducing higher-order goods and services for sale. These will not be required in such great quantities as the lower-order goods and services (everyday requirements), and with their higher threshold values must draw upon much larger areas to produce a reasonable profit. Fig. 22.9 shows the superimposition of the village fields of influence and those at the higher-order market town level.

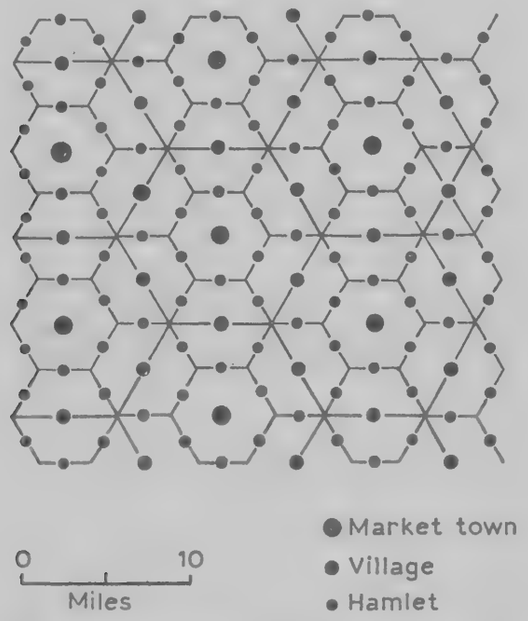


Fig. 22.9 Hamlets, villages and towns in a 'fixed-k' hierarchy where $k = 3$ (Christaller's 'marketing' principle)

As has been seen in this example, the total number of settlements of a certain order served by a central place of the next higher order is three. Christaller termed this the k value. In this case $k = 3$. In Fig. 22.9 a $k = 3$ lattice is shown. However, if the hexagonal net is turned through ninety degrees, the border settlements will be shared by only two central places and thus the k value rises to four, as is shown in Fig. 22.10. Fig. 22.11 shows the process of net enlargement moving one step farther to produce a $k = 7$ lattice. In this latter case it can be seen that no settlements are shared between service areas.

Christaller suggested that in theory once a k value had developed within an area it would apply to all the levels in the hierarchy so formed (see Fig. 22.9 where $k = 3$ at both the village and market town levels). This is Christaller's fixed-k hierarchy principle.

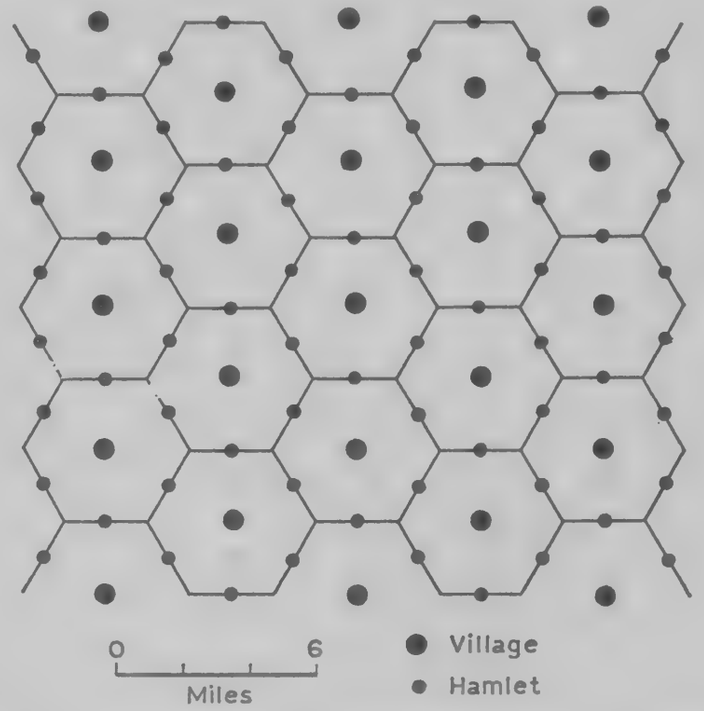


Fig. 22.10 Pattern of service areas of villages in a $k = 4$ hierarchy

Christaller suggested that a fixed k value of 3 would develop where the supply of goods from the central places is to be as near the dependent places as possible. This condition maximises the number of central places in the hypothetical landscape, and is referred to as the marketing principle of Christaller. On the other hand, Christaller suggested that a $k = 4$ system would develop where the cost of constructing transport networks is of importance.

Fig. 22.12 shows the $k = 4$ lattice developed to the order of the market town. Here it can be seen that as many important places as possible lie on any one traffic route. This is Christaller's traffic principle.

If firm administrative control were important, then a fixed-k

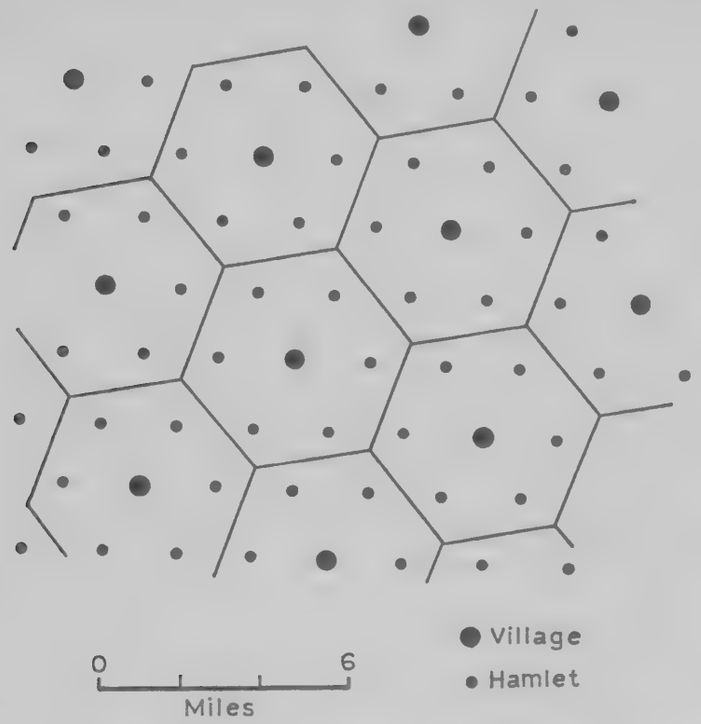


Fig. 22.11 Pattern of service areas of villages in a $k = 7$ hierarchy

system of value 7 would develop. In this case all six dependent places would owe allegiance to the central place only and not have divided loyalties. Fig. 22.13 shows a fixed-k system of value 7 developed through to the market town level, under Christaller's administrative principle.

One of the problems associated with Christaller's fixed-k system of values 3 and 4, is that dependent places are divided in their allegiance to the central places one higher in the hierarchy. Christaller recognised this and overcame it by suggesting that what he referred to as 'nesting' would occur. This is explained in Fig. 22.14. In this diagram, it is suggested that in the case of a

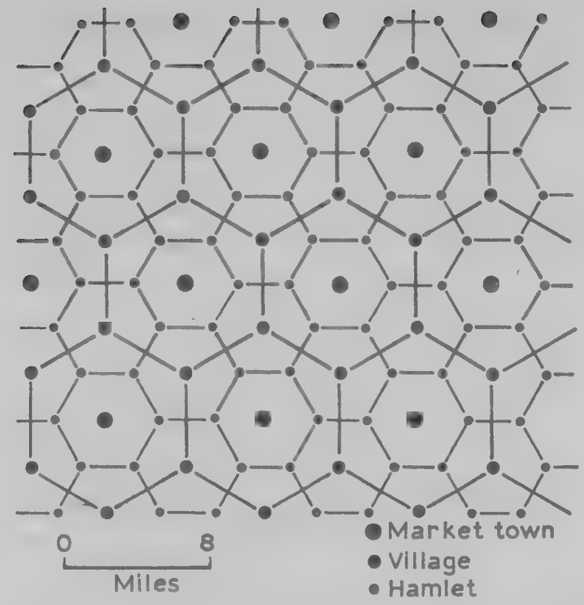


Fig. 22.12 Hamlets, villages and towns in a 'fixed-k' hierarchy where $k = 4$ (Christaller's traffic principle)

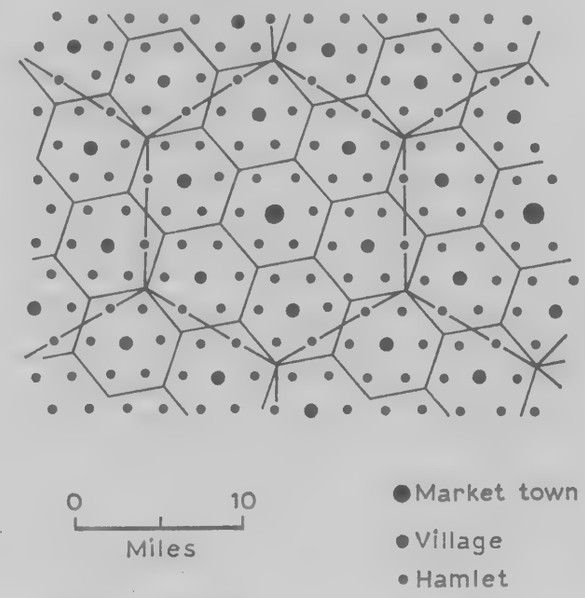


Fig. 22.13 'Fixed-k' hierarchy of $k = 7$ developed to third order of settlement (market towns) (Christaller's administrative principle)

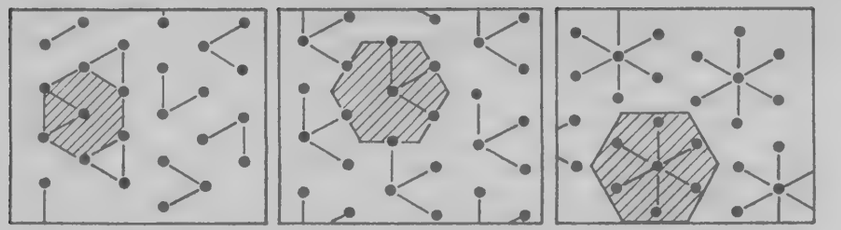


Fig. 22.14 Alternative nesting principles in the Christaller 'fixed-*k*' landscape (after Haggett, *Location Analysis in Human Geography*, 1965)

$k = 3$ hierarchy connections would be set up between two dependent settlements and the central place, and with the $k = 4$ and $k = 7$ hierarchies it is suggested that connections will develop in a similar way in the manner shown.

Note
1. pp. 102–8.

The writings of Everson and FitzGerald have been quoted here because they explain more adequately the Berry and Pred summary of Christaller's hexagonal lattices in relation to his three locating principles. But of the general body of central-place theory considered as a whole, it is worth while making one general point. As Ambrose[1] *has pointed out:*

Implicit in much of this thought is the idea of *spatial competition*. Providers of identical functions are competing for customers. Therefore, since customers are distributed in space, they are in effect competing for space. New functions will tend to locate at points which are as far away as possible from existing functions of the same type. As one provider increases the space he serves, so he decreases the space served by his immediate competitors. As people become more mobile, and as some providers become much larger and more attractive in their prices and range of services than others, then the smaller providers of that function may cease to operate altogether. This situation can be seen quite clearly as increasing numbers of rural dwellers acquire cars which may lead them to transfer some of their custom from the local village shop to a supermarket in some nearby town. (It should be noted, in passing, that there are important exceptions to some of these general statements. Sometimes, particularly within a town, certain functions such as antique shops tend to cluster together rather than space themselves as far apart as possible. The reader may care to theorise about the factors which might bring this situation about.)

Note
1. pp. 126–7.

CENTRAL-PLACE THEORY AS A PREDICTIVE DEVICE

The question arises at this stage as to the value of Christaller's work as a predictive device for indicating the patterns of settlement that might be expected in the real landscape. Again 'Analytical Human Geography'[1] *provides a useful summary:*

We could predict a functional hierarchy of central places since at any given point in time some central places would include functions not yet included by others. Also it would be likely that the larger central places would not be clustered in one part of the region but would be regularly spaced all over it. Additionally we would expect towns of a given importance in the hierarchy to be spaced at a roughly constant distance apart. It also follows, since the number of towns in each size class decreases as one goes up the size hierarchy, that larger towns should tend to be spaced farther apart than smaller towns. To what extent are all these expectations borne out in reality?

The answer is that in many cases, perhaps most, the actual pattern differs considerably from that expected on the basis of the theory. This is partly because central-place theory relates only to one activity of towns; their servicing function to surrounding regions. But, as we know, towns grow for a variety of reasons apart from this. They may grow because they are situated on a resource of some kind which occurs only in a limited area. 'Resource' here might mean a mineral of use for manufacturing (for example, coal), a scenic or climatic factor of use for recreation (for example, a coastline or a reliably warm summer), or a strategic location for transhipment or distribution (for example, a deepwater estuary). All these factors might confer upon the limited area in question

advantages which would lead to a rate of urban growth in excess of that to be expected solely from the demand for central services generated by the immediate tributary region. This helps to explain why two such large cities as Liverpool and Manchester exist within a few miles of each other.

A second factor that distorts the perfect hexagonal scheme is the variations that occur in population density. For the sake of simplicity, we assumed in our example that population was spread all over the region at a constant density. This is unlikely to occur in reality, since factors upon which population density depends, for example soil fertility and climate, vary from place to place. In a relatively densely populated rural area, the threshold population necessary to support a central function occupies a smaller area of land than would be the case in a sparsely populated area. Since it is obviously threshold *population*, not threshold *area*, that determines whether a function will exist, then it is more useful to think in terms of *population space*, not actual space, when analysing central-place systems. Maps drawn in population space are those where given areas on the map are made proportional to the population of those areas on the ground. Such maps are sometimes drawn at election times. The area of each constituency on the map is made proportional to the number of voters it contains and from this map one can tell at a glance the political significance of each constituency. Space distortions of this sort can help us a great deal to improve our comprehension of the spacing and size of service centres.

There is one other important shortcoming of central-place theory that should be considered. The theory includes the assumption that to reach any given function the rural dweller will go always to the nearest possible centre which offers it. In other words it suggests that for each commodity, adjacent tributary regions are mutually exclusive and that there is no zone of overlap in which people go sometimes to one centre and sometimes to another. From our own common knowledge we know this to be untrue, since any one of a large number of reasons may induce us to go farther than we need for a commodity or service. Once we accept that this is true, it becomes less pressing to think in terms of hexagonal tributary areas, since the prime reason for suggesting hexagons was that they fitted together without overlapping.

Note

1. pp. 127–8.

Many other geographers have of course offered criticisms of the Christaller model. Among the best known of these was Edward Ullman. He has commented not only upon the basic theory but on the empirical evidence provided by Christaller from his studies in Germany which is used in support of the model. Ullman, in his paper 'A Theory of Location for Cities' (1941),[1] has written:

Population alone is not a true measure of the central importance of a city; a large mining, industrial or other specialised-function town might have a small tributary area and exercise few central functions. In addition to population, therefore, Christaller uses an index based on number of telephones in proportion to the average number per thousand inhabitants in South Germany, weighted further by the telephone density of the local subregion. A rich area such as the Palatinate supports more telephones in proportion to population than a poor area in the Bavarian Alps; therefore, the same number of telephones in a Palatinate town would not give it the same central significance as in the Alps. He claims that telephones, since they are used for business, are a reliable index of centrality. Such a thesis would not be valid for most of the United States, where telephones are as common in homes as in commercial and professional quarters.

Some better measures of centrality could be devised, even if only the number of out-of-town telephone calls per town. Better still would be some measure of actual central services performed. It would be tedious and difficult to compute the amount, or percentage, of business in each town drawn from outside the city, but some short cuts might be devised. If one knew the average number of customers required to support certain specialised functions in various regions, then the excess of these functions over the normal required for the urban population would be an index of centrality.[2] In several states rural sociologists and others have computed the average number of certain functions for towns of a given size. With one or two exceptions only small towns have been analysed. Retail trade has received most attention, but professional and other services have also been examined.

These studies do not tell us actually what population supports each service, since the services are supported both by town and by surrounding rural population, but they do provide norms of function expectancy which would be just as useful.

A suggestive indicator of centrality is provided by the maps which Dickinson has made for *per capita* wholesale sales of cities in the United States.[3] On this basis centres are distributed rather evenly in accordance with regional population density. Schlier has computed the centrality of cities in Germany on the basis of census returns for 'central' occupations.[4] Refinement of some of our census returns is desirable before this can be done entirely satisfactorily in the United States, but the method is probably the most promising in prospect.

Another measure of centrality would be the number of automobiles entering a town, making sure that suburban movements were not included. Figures could be secured if the state-wide highway planning surveys in forty-six states were extended to gather such statistics.

The central-place scheme may be distorted by local factors, primarily industrial concentration or main transport routes. Christaller notes that transportation is not an areally operating principle, as the supplying of central goods implies, but is a linearly working factor. In many cases central places are strung at short intervals along an important transport route, and their tributary areas do not approximate the ideal circular or hexagonal shape but are elongated at right-angles to the main transport line. In some areas the reverse of this normal expectancy is true. In most of Illinois, maps depicting tributary areas show them to be elongated parallel to the main transport routes, not at right-angles to them. The combination of nearly uniform land and competitive railways peculiar to the state results in main railways running nearly parallel and close to one another between major centres.

In highly industrialised areas the central-place scheme is generally so distorted by industrial concentration in response to resources and transportation that it may be said to have little significance as an explanation for urban location and distribution, although some features of a central-place scheme may be present, as in the case of Cologne and the Ruhr.

In addition to distortion, the type of scheme prevailing in various regions is susceptible to many influences. Productivity of the soil,[5] type of agriculture and intensity of cultivation, topography, governmental organisation, are all obvious modifiers. In the United States, for example, what is the effect on distribution of settlements caused by the sectional layout of the land and the regular size of counties in many states? In parts of Latin America many centres are known as 'Sunday towns'; their chief functions appear to be purely social, to act as religious and recreational centres for holidays – hence the name 'Sunday town'. Here social rather than economic services are the primary support of towns, and we should accordingly expect a system of central places with fewer and smaller centres, because fewer functions are performed and people can travel farther more readily than commodities. These underlying differences do not destroy the value of the theory; rather they provide variations of interest to study for themselves and for purposes of comparison with other regions.

The system of central places is not static or fixed; rather it is subject to change and development with changing conditions.[6] Improvements in transportation have had noticeable effects. The provision of good automobile roads alters buying and marketing practices, appears to make the smallest centres smaller and the larger centres larger and generally alters trade areas.[7] Since good roads are spread more uniformly over the land than railways, their provision seems to make the distribution of centres correspond more closely to the normal scheme.[8]

Christaller may be guilty of claiming too great an application of his scheme. His criteria for determining typical-size settlements and their normal number apparently do not fit actual frequency counts of settlements in many almost uniform regions as well as some less rigidly deductive norms.[9]

Bobeck in a later article claims that Christaller's proof is unsatisfactory.[10] He states that two-thirds of the population of Germany and England live in cities and that only one-third of these cities in Germany are real central places. The bulk are primarily industrial towns or villages inhabited solely by farmers. He also declares that exceptions in the rest of the world are common, such as the purely rural districts of the Tonkin Delta of Indochina, cities based on energetic entrepreneurial activity, as some Italian cities, and world commercial ports such as London, Rotterdam and Singapore. Many of these objections are valid; one wishes that Christaller had better quantitative data and were less vague in

places. Bobeck admits, however, that the central-place theory has value and applies in some areas.

The central-place theory probably provides as valid an interpretation of settlement distribution over the land as the concentric-zone theory does for land use within cities. Neither theory is to be thought of as a rigid framework fitting all location facts at a given moment. Some, expecting too much, would jettison the concentric-zone theory; others, realising that it is an investigative hypothesis of merit, regard it as a useful tool for comparative analysis.

Notes

1. *American Journal of Sociology*, vol. 46, pp. 853–64.
2. In Iowa, e.g., almost all towns of more than 450 inhabitants have banks, half of the towns of 250–300 and 20 per cent of the towns of 100–150.
3. R. E. Dickinson, 'The Metropolitan Cities of the United States', *Geogr. Rev.* vol. 24 (1934) 280–1.
4. O. Schlier, 'Die Zentralen Orte des Deutsches Reichs', *Zeitschrift der Gesellschaft für Erdkunde X zu Berlin* (1951) 161–70.
5. Cf. the emphasis of Sombart, Adam Smith and other economists on the necessity of surplus produce of land in order to support cities. Fertile land ordinarily produces more surplus and consequently more urban population, although 'the town . . . may not always derive its whole subsistence from the country in its neighbourhood . . .'.
6. The effects of booms, droughts and other factors on trade-centre distribution by decades are brought out in Landis's studies for South Dakota and Washington. Zimmerman and Smith also show the changing character of trade-centre distribution. Melvin calls attention to a 'village population shift lag'; in periods of depressed agriculture villages in New York declined in population approximately a decade after the surrounding rural population had decreased.
7. Most studies indicate that only the very smallest hamlets (under 250 population) and crossroads stores have declined in size or number. The larger small places have held their own. Zimmerman in 1930 notes that crossroads stores are disappearing and are being replaced by small villages. 'Farm Trade Centers in Minnesota 1905 to 1929', *University of Minnesota Agricultural Experimental Station Bulletin*, vol. 269, p. 41. He states further: 'It is evident that claims of substantial correlation between the appearance and growth of the larger trading centre and the disappearance of the primary centre are more or less unfounded. Although there are minor relationships, the main change has been a division of labour between the two types of centres rather than the complete obliteration of the smaller in favour of the larger.'
8. Smith states: 'There has been a tendency for centres of various sizes to distribute themselves more uniformly with regard to the area, population and resources of the state. Or the changes seem to be in the direction of a more efficient pattern of rural organisation. This redistribution of centres in conjunction with improved methods of communication and transportation has placed each family in frequent contact with several trade centres. . . .' 'Farm Trade Centers in Louisiana 1905 to 1931', *Louisiana State University Bulletin* (1933) 54.

   In contrast, Melvin, writing about New York State before the automobile had had much effect, states *Rural Population of New York, 1855–1925*, p. 90: 'In 1870 the villages . . . were rather evenly scattered over the entire state where they had been located earlier in response to particular local needs. By 1920, however, the villages had become distributed more along routes of travel and transportation and in the vicinity of cities.'
9. This statement is made on the basis of frequency counts by the author for several mid-western states.
10. H. Bobek, 'Übereinige functionelle Stadt-typen und ihre Beziehungen zum Lande', *Comptes Rendus du Congrès International de Géographie* (Amsterdam, 1938).

ANALYSES OF HIERARCHIES

Ullman wrote his critique of Christaller in 1941. More recently, other and possibly more sophisticated analyses of central-place theory have appeared, mainly aimed at investigating further the notion of hierarchies of settlement by size and by function in terms of goods and services supplied. As an explanation to the background to one representative paper in this field, Ambrose[1] *has written:*

If classical central-place theory were exactly correct, central places, when ranked in order of size, would produce a 'stepped' size hierarchy, not a continuous one. In other words, all towns would fall into one of a limited number of size classes. A simplified explanation of this expectation is that the hexagonal service areas are all of one of a limited number of sizes because they are all based on a set of fundamental hexagons of equal size. If, as pure theory sets out, there is a constant relationship between the size of a service area and its central place, then all the central places must be of a limited number of sizes and the population size of centres, when they are ranked in size order, will not vary continuously from, say, one hundred to one million. In reality, towns of all population sizes appear to exist, although we do speak instinctively of hamlets, villages, towns and cities as if settlements of four size classes could be differentiated. It is curious how often, when the population size of each town in a region or country is graphed in size order (this is called a rank-size graph), 'plateaux' appear on the curve. This indicates that towns of certain population size ranges

are comparatively frequent and those of other sizes rare or non-existent.

Note

1. *Analytical Human Geography*, p. 130.

All the same A. Lösch,[1] a German geographer, could not accept a notion which suggested the tendency for 'grouping' in a rank-size hierarchy as the Christaller model does. He therefore rejected the idea of a fixed-k system discussed earlier by Everson and FitzGerald, since he did not think it approached reality. Instead, he took the original model and developed it in a more elaborate form. He did this by superimposing all the various hexagonal systems mentioned in relation to the Christaller system ($k = 3$, $k = 4$ and $k = 7$) besides many more on one point. Then by rotating them he was able to identify six sectors with many relatively high-order settlements (which he described as 'city-rich' sectors) and six sectors with relatively few high-order settlements ('city-poor' sectors).

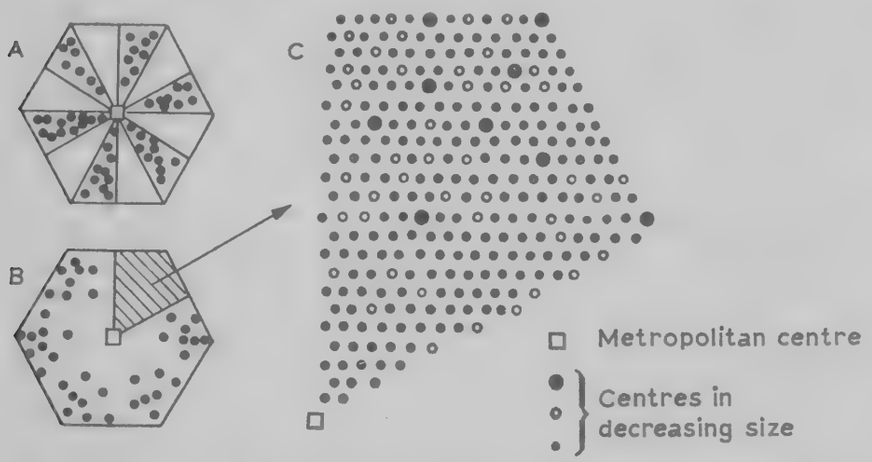


Fig. 22.15 Löschian landscape developed by the rotation of a number of *k*-system hexagons, developing 'city-rich' and 'city-poor' sectors (after Haggett, *Location Analysis in Human Geography*, 1965):
(*a*) 'City-rich' and 'city-poor' sectors in Löschian landscape
(*b*) Distribution of large cities in Löschian landscape
(*c*) Distribution of all centres in one sector
Source: Everson and FitzGerald, *Settlement Patterns*, p. 109

Using this method of a variable-k hierarchy, Lösch considered that the pattern of settlements produced was much closer to a real-world situation. It produced, in contrast to Christaller's settlements in tiers, a continuous sequence of settlement size. He also said that his idea applied not only to population sizes of settlements, but also 'to the function-content in terms of supplies of goods and services'.[2]

This problem of the continuous or stepped hierarchy has exercised other research workers. B. J. L. Berry and W. L. Garrison were concerned to discover whether the thirty-three central places in a county in Washington State, U.S.A., fitted into a stepped or grouped hierarchy or whether they were in the form of a continuous hierarchical curve. In order to carry this work through, Berry and Garrison analysed the hierarchy of functions provided by the towns to discover if the centres fell into groupings according to functional significance.

Their conclusions were threefold. First, there seemed to be a rough positive relationship between population size and functional importance. Second, four centres were very poor in service provision in relation to their size. Third, three reasonably distinct groups emerged when the towns were analysed according to total number of activities or functions. These groupings are shown by vertical broken lines in Fig. 22.17 of their article 'The Functional Bases of the Central Place Hierarchy' (1958)[3] which follows. The groupings, lettered A, B and C, provide the bases for the grouping of centres under the same letters in Tables 22.2 and 22.3.

Berry and Garrison's paper is a relatively short one and is therefore now worth quoting in full. However, it should be noted that it contains a number of technical terms used by social scientists. These can easily be understood with the help of a dictionary. Also, although a basic knowledge of standard statistical techniques would be helpful in fully understanding the methodology of the paper, the general argument is still comprehensible if these are ignored.

The paper begins with introductory comments:

It is obvious that urban centres differ, each from others. On the intuitive level one notion of difference is that of classes of urban centres. The wealth of descriptive terms available illustrates this notion: hamlet, village, town, city, metropolis and the like. The present study is concerned with this problem of the differentiation of centres into broad classes. In particular it provides original and urgent evidence that a system of urban centre classes exists of the type identified on an intuitive level above.

There are several reasons for producing evidence of a system of classes (hereafter termed the hierarchical class-system) at this time. A considerable body of theory relating to city size, function

and arrangement has accumulated. One of the implications of this theory is that there exists a hierarchical class-system. Ample evidence is available that other implications of this theory are valid, namely, that larger centres are functionally more complex than smaller centres, with this increasing functional complexity being accompanied by increasing size of the urban complementary region, and that by virtue of the differential provision of central functions there is interdependence between urban centres in the provision of central goods and services. On the other hand, there has been no satisfactory evidence provided that would suggest that a hierarchical class-system of centres does indeed exist.

Despite many attempts at the assignment of towns to classes or grades, the converse has seemed to be more likely. The *a priori* methods used by most of the studies have led to serious doubts as to whether a class-system is present in other than arbitrary form, and to the alternate idea that perhaps instead of a class-system there only exists differentiation along a continuum. Vining, for example, has written about the single best-known empirical study of the central-place hierarchy, that of John E. Brush:[4]

> There is no evidence . . . that exactly three natural partitions may be observed in this array of numbers of establishments. . . . The terms hamlet, village and town are convenient modes of expression; but they do not refer to structurally distinct natural entities. . . . Clearly, it is arbitrary to divide the array into three partitions rather than into a greater or lesser number; and similarly arbitrary is the determination of where to put the dividing points separating the different classes or types. Having drawn the lines, one may list certain kinds of activities which are typically found within each of the designated classes of centre, . . . not all members of a class will contain all the activities listed and most of the communities within a class will contain activities not listed. [This] . . . is not an independently derived basis for a classification of communities by type. Rather it is itself derived from a previous partitioning of an array which appears as something similar to an arrangement of observations that have been made upon a continuous variable.[5]

These criticisms of Brush's study may well be valid. That he used an arbitrary division and then proved what he had in fact assumed is without question (this criticism applies to more studies of central places than is generally realised). The ensuing discussion serves to disprove the implication which Vining drew from this fact, that central places are differentiated along a continuum rather than in a class-system.

The finding of a hierarchical class-system verifies implications of theory and thus performs one classic function of empirical research in the accretion of knowledge. In the present case the need for the performance of this classic function is quite pressing because alternate theories are available to explain characteristics of systems of cities. To the extent that verification is provided for previously unverified schemes, the present study increases the attractiveness of these schemes versus more recent schemes which do not include hierarchical concepts. It might be added that the present study provides examples of taxonomic methods for the empirical study of the hierarchical class-system and these methods may be found useful in ensuing studies; it was designed as a *critical study* which may be checked in the study area (Snohomish County, Washington) and which is capable of being reproduced by the methods provided in other areas and at levels of the central-place hierarchy other than the universe of small centres considered in the present case.

THE HIERARCHICAL CONCEPT

The hierarchical class-system implication is an integral part of the spatial model of central places developed by Walter Christaller, the generic base and single most important statement of central-place theory. The model states that central places belong to one or another of class subsets. Each class possesses specific groups of central functions and is characterised by a discrete population level of its centres. Note that (1) classes are arranged one to another in a hierarchy such that the central places of functionally more complex classes possess all the groups of functions of less complex classes plus a group of functions differentiating them from the central places of less complex classes, and (2) discrete population levels of the central places of each class are thought to arise because that income which supports the population of a central place is brought into the centre by the activities which provide goods and services for surrounding consumers. Since the central places of each class possess discrete groups of activities, they also tend to have discrete population levels.

AN OBSERVED HIERARCHY OF CENTRAL PLACES

Evidence is provided here that a hierarchy of central places does exist. In Snohomish County, Washington, the area in which the present study was undertaken, it was found that centres could be arranged into three types, called A, B and C (Table 22.1). These

TABLE 22.1

CLASSES OF CENTRAL PLACES AND ASSOCIATED GROUPINGS OF CENTRAL FUNCTIONS IN SNOHOMISH COUNTY, WASHINGTON

Central functions	*Classes of central places* A*	B	C
Variates:			
Group 1_1	*p*†	*f*	*f*
Group 2_1		*p*	*f*
Group 3_1		*s*	*f*
Attributes:			
General store	*p*		
Group 1_2	*s*	*p*	*f*
Group 2_2		*p*	*f*
Group 3_2		*s*	*p*

* A Class A centre, for example, provides a partial range of functions of group 1_1, a few of the functions of group 1_2, and most, although not all, of the centres will have general stores.

† *p* = partial range; *f* = full range; *s* = some.

three types were defined on the basis of the presence of urban functions 1_1, 1_2, . . . , etc., in varying degrees. In other words, in Snohomish County it was found to be possible to treat central places taxonomically and derive centre types A, B and C (say, hamlets, villages and towns, respectively) on a functional basis. It was shown that these centre types differ more one type from another than they differ within types.

The data utilised involved thirty-three of the smaller central places of the county (Fig. 22.1*b*). Since the present study was in the interest of expediency limited to small central places, excluded from the field survey was the largest centre in the county, Everett. Several of the smaller communities on the northern and southern margins of the county were also excluded. Reasons for the exclusion of these centres, an elaboration of the data used in the study, and relevant maps of the travel patterns of the rural residents to the central places are available elsewhere.

Of the thirty-three centres studied, twenty were of Class A, which contains the lowest complex of central functions; nine were of Class B; and four were of Class C, which contains the most complex grouping of functions. In these centres considered *in toto* it was possible to recognise sixty-three kinds of central functions (in addition four non-central functions and other characteristics of the centres were included within the study).[6] These could be differentiated into seven classes of functions (three and four classes in each of the broad groups, respectively), distinguished in terms of the hierarchical concept.

This preliminary summary of the data is at best quite gross. Any discussion of the exact nature of the hierarchy requires an elaboration of the functional classes and the classes of centres, and a clarification of what is meant by 'some', a 'partial range' and a 'full range' of central functions. It is necessary also to detail the evidence on which the study is based, both to verify the viability of the classes and, because the classes evolve directly from the evidence, to establish the characteristics of the classification. The ensuing discussion provides this elaboration.

It will be noted that in the above paragraphs the hierarchical class-system of centres was assumed to follow from the class-system of functions. The assumption is derived from the implications of theory. A further note on this notion is included in the ensuing parts of the study.

THE ANALYSIS

The problem before us was to determine whether central functions fall into groups of classes and, if so, whether these classes were associated with classes of central places, as theory suggests they ought to be. Therefore, the initial step in the research was to rank both central functions and central places. The second step was to apply tests to determine whether groupings occurred and, if so, to discover whether there were significant differences between the groupings isolated.

RANKING THE CENTRAL FUNCTIONS

Fifty-two of the central functions were variates – that is, numbers of stores performing these functions varied from place to place. The

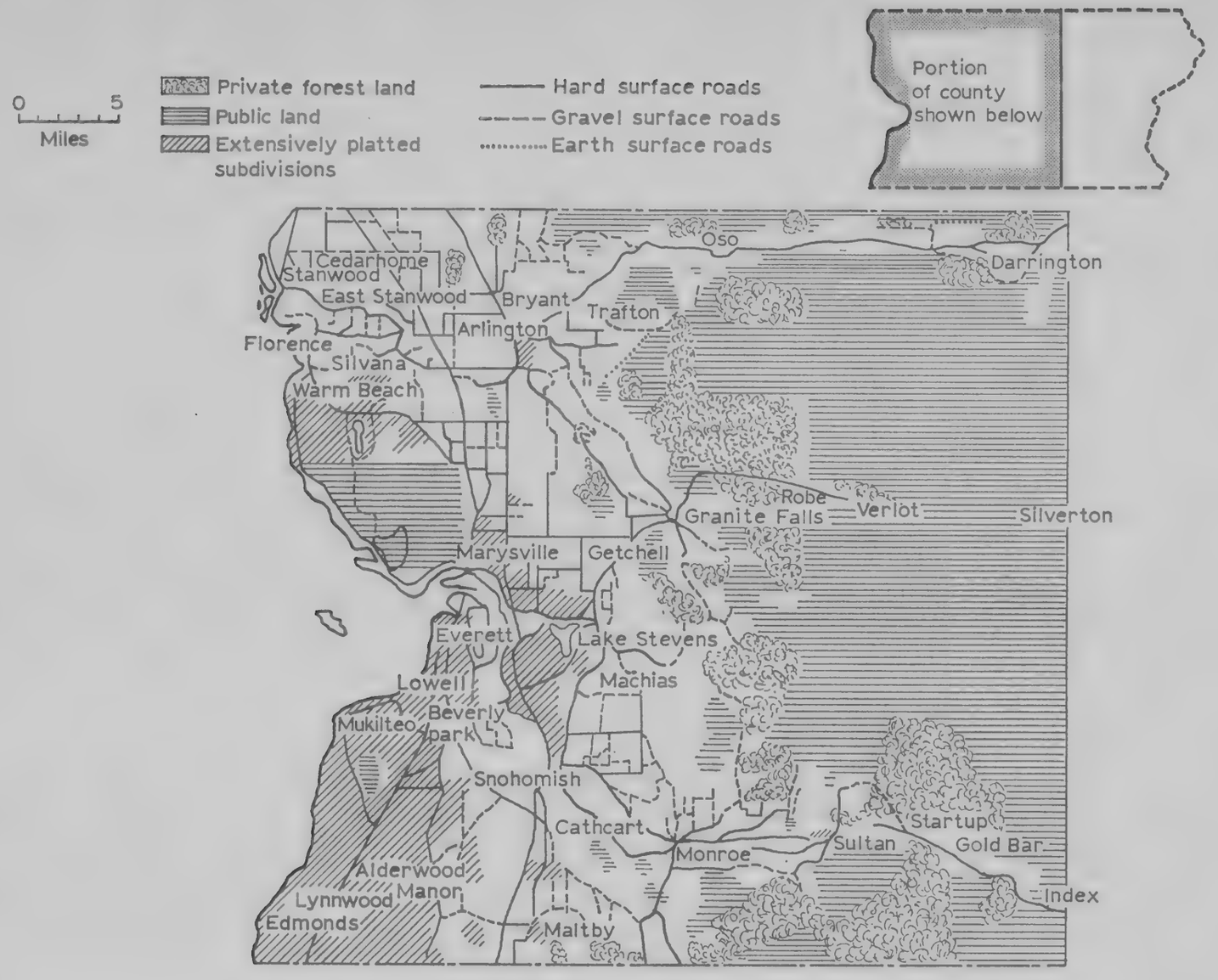


Fig. 22.16 Locational features of the study area, Snohomish County, Washington

TABLE 22.2

THE FUNCTIONAL BASES OF CLASSES OF CENTRAL PLACES IN SNOHOMISH COUNTY: VARIATES

	Central functions	*Threshold population*	*B-value*	*Central Places*																																		
				C				*B*									*A*																						
				Marysville	Snohomish	Arlington	Monroe	Edmonds	Stanwood	East Stanwood	Lynnwood	Sultan	Mukilteo	Darrington	Granite Falls	Alderwood Manor	Lake Stevens	Beverly Park	Lowell	Silvana	Gold Bar	Startup	Warm Beach	Index	Machias	Cathcart	Maltby	Oso	Cedarhome	Bryant	Robe	Getchell	Florence	Trafton	Silverton	Verlot
	Population	...	...	2460	3494	1915	1684	2996	720	390	500	850	900	974	600	600	2586	725	1600	300	325	300	314	220	200	175	700	200	100	150	50	25	300	25	15	20
	Number of activities	...	...	64	62	59	56	42	42	38	36	34	28	25	25	22	16	13	12	10	9	8	6	6	5	5	4	4	4	3	3	3	2	2	2	1
I_1	Filling stations	196	1·35	9	3	6	8	7	4	3	9	6	3	3	4	4	3	1	1	1	1	2	1	...	2	1	2	2	1	1	1	1	1	...	1	...
	Food stores	254	1·74	5	6	4	6	6	2	2	3	3	1	2	3	3	2	1	1	2	2	2	...	...	...	...	...	...	...	1	...	...	...	...	...	...
	Churches	265	1·29	5	17	9	8	8	3	3	3	4	2	2	2	2	2	...	1	1	...	2	1	...	1	...	1	...	1	...	...	...	...	...	...	...
	Restaurants	276	1·33	14	7	5	6	5	2	3	6	3	1	1	2	1	1	...	2	...	4	1	...	1	...	...	...	...	...	...	1	...	...	...	...	...
	Taverns	282	1·65	2	5	3	5	5	2	5	3	3	1	3	2	...	2	1	1	1	1	2	...	1	1	...	...	...	...	...	1	...	...	...	...	...
	Elementary schools	322	1·67	6	6	4	2	1	3	1	1	1	2	1	1	1	1	...	....	1	1	1	1	1	1	1	1	1	...	...	1	...	1	...	...	...
	Physicians	380	1·42	3	5	10	3	5	2	1	1	1	...	...	1	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Real-estate agencies	384	1·40	6	7	3	2	....	2	2	6	...	1	...	...	2	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Appliance stores	385	1·46	2	10	4	1	4	1	2	1	1	...	1	1	1	....	..	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Barber shops	386	2·39	4	4	4	3	2	2	1	1	1	1	1	1	1	2	1	1	...	1	...	...	...	...	1	...	...	...	...	...	...	...	...	...	...	...
	Auto dealers	398	1·35	2	3	10	4	2	1	4	2	...	...	...	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Insurance agencies	409	1·32	....	10	1	4	....	2	3	1	1	...	...	1	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Bulk oil distributors	419	1·56	6	5	5	2	....	3	1	...	2	...	1	1	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Dentists	426	1·57	1	5	4	3	4	1	1	5	...	...	...	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Motels	430	1·56	6	.....	.....	2	4	...	...	3	2	...	3	...	...	1	...	...	...	2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Hardware stores	431	1·90	2	6	3	3	3	...	2	1	1	...	1	1	2	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Auto repair shops	435	1·72	4	6	1	....	1	2	1	1	...	...	...	...	...	....	1	1	...	...	...	...	...	...	1	...	...	...	...	1	...	...	...	...	...
	Fuel dealers (coal, etc.)	453	1·78	4	2	4	1	5	1	...	3	...	...	2	...	1	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Drug stores	458	2·23	3	3	2	2	3	1	1	1	1	1	1	1	1	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Beauticians	480	1·89	5	1	1	3	4	...	1	...	1	1	...	...	2	....	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
B_1	Auto parts dealers	488	1·94	3	3	4	1	2	2	2	2	...	...	1	...	...	....	...	...	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Meeting halls	525	2·01	1	3	1	4	....	...	1	1	1	1	1	1	1	....	...	...	...	...	...	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...
	Animal feed stores	526	1·79	4	2	2	4	....	...	...	...	...	...	...	...	1	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Lawyers	528	2·12	1	3	2	4	2	2	2	1	1	...	...	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Furniture stores, etc.	546	1·85	2	1	1	2	2	...	1	4	...	...	...	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Variety stores: 5 & 10	549	2·30	2	3	2	1	4	1	2	1	...	...	...	...	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Freight lines and storage	567	2·04	4	3	2	3	....	1	1	...	...	...	2	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Veterinaries	579	1·97	1	1	3	1	....	1	2	...	...	...	...	...	1	....	...	...	...	...	...	...	...	...	...	...	1	...	...	...	...	...	...	...	...	...
	Apparel stores	590	2·53	3	3	1	2	2	1	2	...	1	1	1	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Lumber yards and woodworking	598	2·49	1	3	1	1	1	...	1	3	1	1	1	...	...	....	...	1	...	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Banks	610	2·05	1	1	2	1	2	2	1	1	1	...	1	1	...	1	...	....	...	...	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Farm implement dealers	650	1·95	....	1	....	1	....	2	...	2	...	...	...	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Electric repair shops	693	2·62	1	1	2	3	1	1	1	1	...	...	...	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	Florists	729	2·40	1	2	2	3	....	...	...	...	...	...	...	...	...	....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

TABLE 22.2—*continued*

	Central functions	*Threshold population*	*B-value*	*Central places*																																				
				C				*B*										*A*																								
				Marysville	Snohomish	Arlington	Monroe	Edmonds	Stanwood	East Stanwood	Lynnwood	Sultan	Mukilteo	Darrington	Granite Falls	Alderwood Manor	Lake Stevens	Beverly Park	Lowell	Silvana	Gold Bar	Startup	Warm Beach	Index	Machias	Cathcart	Maltby	Oso	Cedarhome	Bryant	Robe	Getchell	Florence	Trafton	Silverton	Verlot				
	Population	....	....	2460	3494	1915	1684	2996	720	390	500	850	900	974	600	600	2586	725	1600	300	325	300	314	220	200	175	700	200	100	150	50	25	300	25	15	20				
	Number of activities	...	....	64	62	59	56	42	42	38	36	34	28	25	25	22	16	13	12	10	9	8	6	6	5	5	4	4	4	3	3	3	2	2	2	1				
2_1	High schools	732	3·64	2	2	1	2	1	1	1	...	2	1	1	1	...	1	...																						
	Dry cleaners	754	3·56	2	1	2	1	3	...	...	1	...	...	1	1	...	....	1																						
	Local taxi services	762	2·89	1	1	1	2	2	1	1	2	...	...	...	...	...	....	....																						
	Billiard hall and bowling alleys	789	2·56	1	3	2	....	....	...	...	...	...	1	...	...	...	....	...																						
	Jewellery stores	827	3·26	1	2	2	2	1	...	1	1	1	...	...	...	...	....	...																						
	Hotels	846	2·92	1	1	2	....	....	1	...	...	...	2	...	...	...	....	...																						
	Shoe repair shops	896	2·45	....	....	1	1	3	1	1	1	1	...	...	1	...	....	...																						
	Sporting goods stores	928	3·30	2	2	....	1	....	1	...	...	1	1	...	...	...	....	...																						
	Frozen food lookers	938	2·99	1	....	2	1	....	...	...	...	...	...	...	...	1	....	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...				
	Sheet metal workers	1076	3·80	2	1	1	1	....	...	...	...	...	...	...	1	1	....	...																						
	Department stores	1083	4·50	1	2	2	1	1	1	...	...	...	1	...	...	...	1	...																						
	Optometrists	1140	3·88	1	1	1	2	....	1	...	...	...	...	...	...	...	....	...																						
	Hospital and clinics	1159	3·77	1	1	2	....	....	...	...	...	...	1	...	...	...	....	...																						
	Undertakers	1214	4·19	1	1	1	....	2	1	1	...	...	...	...	...	...	....	...																						
3_1	Photographers	1243	4·56	1	2	1	....	....	...	...	...	...	1	...	...	...	....	...																						
	Public accountants	1300	4·11	1	2	....	1	....	...	...	...	...	...	...	...	...	....	...																						
	Laundries and laundromats	1307	4·66	2	....	....	....	1	...	...	...	...	1	...	...	...	....	1																						
	Health practitioners	1424	4·55	1	2	....	....	1	...	...	...	...	...	...	...	...	....	...																						

remaining functions and other characteristics considered in the study were attributes – that is, central places either possessed a unit performing this function or did not possess such a unit. The method of ranking varied between these two sets.

Variates

Previous empirical studies indicated that relationships exist between the population of a central place and the number of units of any function which that place possesses and this is, of course, clear from common knowledge. Christaller suggested that through the working of the income mechanism the population of a centre was a function of the number of types of central goods and services the central place provided. Hence it was specified that the population of a centre is a function of the number of stores of each type. Fifty-two scatter diagrams were prepared with population, P, and number of stores, N, as parameters to determine the relationships between P and N for each function. Each of the diagrams had thirty-three points, one for each of the thirty-three central places. Best-fitting curves of the exponential growth series $P = A\ (B^N)$, where A and B are the parameters to be estimated, were fitted to each of the scatters using standard least squares techniques, after logarithmic conversion.

TABLE 22.3

THE FUNCTIONAL BASES OF CLASSES OF CENTRAL PLACES IN SNOHOMISH COUNTY: ATTRIBUTES

	Central places	*Central functions*														
			1_2			2_2									3_2		
		General store	Feed mill	Post office	Weekly newspaper	Telephone exchange	Incorporated city	Movie theatre	Bakery	Electricity distribution	Water supply system	Sewage system	State liquor store	Public library	Printing press	Used furniture dealer
	r_{pb}	− 0·347	0·265	0·29	0·458	0·584	0·615	0·616	0·619	0·646	0·646	0·691	0·691	0·758	0·759	.....
C	Marysville	.....	1	1	1	1	1	1	1	1	1	1	1	1	1	.....
C	Snohomish	.....	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	Arlington	.....	1	1	1	1	1	1	1	1	1	1	1	1	1	.....
C	Monroe	.....	1	1	1	1	1	1	1	1	1	1	1	1	.....	.....
B	Edmonds	.....	.....	.....	.....	.....	1	.....	1	1	1	1	1	1	1	.....
B	Stanwood	.....	.....	.....	1	1	1	1	.....	1	1	1	.....	1	.....	.....
B	East Stanwood	.....	.....	1	.....	.....	1	.....	1	.....	.....	.....	1	.....	.....	.....
B	Lynnwood	.....	.....	1	.....	.....	.....	.....	1	1	1	1	.....	.....	.....	.....
B	Sultan	.....	1	1	1	1	1	.....	.....	1	1	.....	1	1	.....	.....
B	Mukilteo	1	.....	1	.....	.....	1	.....	1	1	1	1	.....	.....	.....	.....
B	Darrington	.....	.....	1	.....	1	1	1	.....	.....	.....	.....	.....	.....	.....	.....
B	Granite Falls	.....	.....	.....	1	.....	1	.....	.....	1	1	.....	.....	.....	.....	.....
B	Alderwood Manor	.....	1	1	.....	.....	.....	.....	1	.....	.....	.....	.....	1	.....	.....
A	Lake Stevens	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	1	.....	.....
A	Beverly Park	.....	.....	1	.....	.....	.....	.....	.....	1	1	1	.....	.....	.....	.....
A	Lowell	.....	.....	1	.....	.....	.....	.....	.....	1	1	1	.....	.....	.....	.....
A	Silvana	1	1	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Gold Bar	.....	.....	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Startup	.....	.....	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Warm Beach	1	.....	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Index	1	.....	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Machias	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Cathcart	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Maltby	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Oso	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Cedarhome	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Bryant	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Robe	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Getchell	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Florence	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Trafton	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Silverton	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
A	Verlot	1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

Given these 52 best relationships it was then possible to rank the central functions on the basis of the threshold population of the centre which was necessary for the first complete store to appear, that is, by the value of P where $N = 1$ (Table 22.2).

Attributes

Relationships between the fifteen attributes and the populations of the centres in which they appear were determined by calculating the point biserial coefficient of correlation between each of the activities and the population of the centres. Activities were then ranked in ascending order of these coefficients, r_{pb}, since it was observed that higher correlations were associated with occurrence in larger centres (Table 22.3).[7]

THE MEASUREMENT OF GROUPINGS BETWEEN RANKED ACTIVITIES

Given the ranks of central functions it was then possible to employ tests to determine whether there were associated groups among them. The tests used again varied between the two sets of central functions.

Variates

Threshold sizes were treated as points upon the population-size continuum. Using the techniques developed by P. J. Clark, this distribution of points was tested for randomness by using a χ^2 (chi-square) test for significant differences between expected and observed reflexive relationships.[8] The tests showed that the observed distribution of points was non-random at the 0·05 level of significance, and this observed distribution was non-random in a grouped rather than a 'more even than random' manner.

Given this tendency for there to be grouping, and using Clark's criterion of a group – that every member of the group should be closer to some other member of the group than to any other point – three groups of central functions, 1_1, 2_1 and 3_1, were found to be present (Table 22.2).

Attributes

Tests of significant differences between the r_{pb}'s were made at the 0·05 level of significance using standard techniques, and twelve of the attributes were found to fall into three groupings, 1_2 and 2_2 and 3_2, with general stores existing as a class apart (Table 22.3).

RANKING AND GROUPING THE CENTRES

The central places were represented as points on the continuum of functional complexity by the number of functions each possessed. Clark's test of randomness was then applied to this distribution

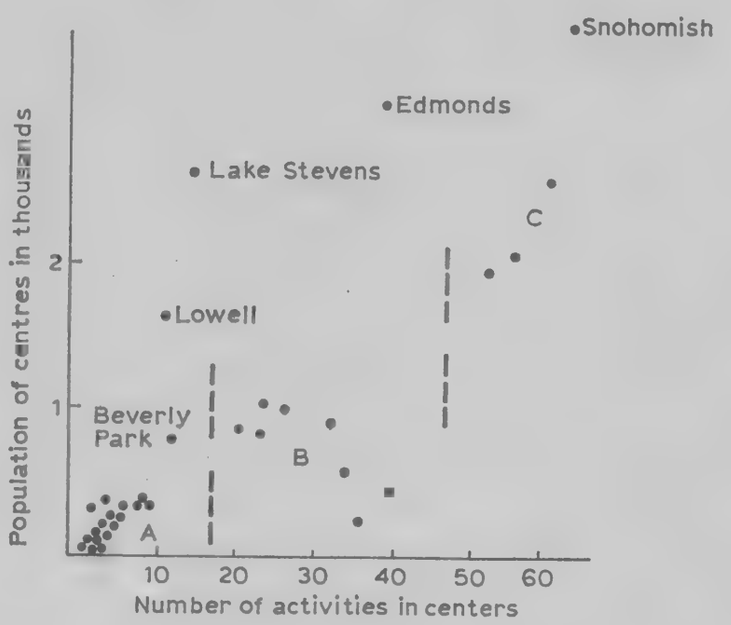


Fig. 22.17 Classes of central places in Snohomish County

of points and the chi-square test showed that the distribution of points was non-random and grouped. Using Clark's concept of a group, three groups of central places, here termed A, B and C, could be distinguished (Fig. 22.17 and Tables 22.2 and 22.3).

THE CLASSES OF CENTRAL PLACES

Given groups of central places defined on the basis of numbers of functions performed, and groups of functions, it remained to be determined whether the latter could be said to be significantly associated with the former. Tables 22.2 and 22.3 were constructed and, utilising the groupings, Table 22.4 derived. This table records the number of stores per function per central place.

TABLE 22.4

NUMBER OF STORES PER FUNCTION PER CENTRAL PLACE IN THE CLASSES OF FUNCTIONS AND OF CENTRAL PLACES IN SNOHOMISH COUNTY, WASHINGTON

Central functions	*Classes of central places* A	B	C
Variates:			
Group 1_1	0·65*	2·91	6·29
Group 2_1	0·04	0·77	2·65
Group 3_1	0·01	0·21	1·00
Attributes:			
General store	0·70	0·11	0·00
Group 1_2	0·20	0·45	1·00
Group 2_2	0·04	0·50	1·00
Group 3_2	0·03	0·28	0·88

* A value of 1·0 means that every centre of the particular class in question will tend to have one store providing each function of the group of functions in question (for example, Class C centres with functions of group 2_2).

Analysis of variance between the cells in Table 22.4 demonstrated significant differences to exist both between groups and between centres at the 0·95 level of confidence. Therefore, variations between groups are greater than variations within groups. This confirms the differential nature of the groupings derived utilising the Clark techniques and reveals the nature of the hierarchy of central places. As an extra step, it was found that general stores, the attribute existing as a group apart, were complementary with food stores (part of group 1_1), the coefficient of correlation being −0·634.

FINDINGS

The data developed in constructing Tables 22.2, 22.3 and 22.4 can be generalised into the gross findings which were presented earlier in Table 22.1. It is apparent how the notion of classes of central places is defined with respect to the notion of functions. What is more, a technique has been provided whereby the hierarchical system may be isolated and identified, and the exact identification of an hierarchical system has been completed for Snohomish County. This is readily comprehended upon close examination of Tables 22.2 and 22.3.

It may also be shown that the three classes of towns tend towards discrete levels of population, as suggested by Christaller (Fig. 22.17). There are only four exceptions to the generalisation of levels in terms of population – Beverly Park, Lowell, Lake Stevens and Edmonds. It is significant that these are centres which have experienced very rapid increases in population in recent years, becoming dormitories for the Seattle area. Their functional paucity may be attributed (*a*) to the time-lag existing between population growth and the development of service industries, and (*b*) to the fact that, compared to other centres of their size, they serve relatively few people residing in immediately surrounding areas. As the time-lag diminishes, the centres can be expected to experience increasing functional complexity and possibly a rise in class status. But the dominant dormitory function suggests that some differential will persist *vis-à-vis* centres of comparable size in the county.

Notes

1. *The Economics of Location* (New Haven, 1954).
2. Op. cit., p. 108.
3. *Economic Geography*, vol. 34, pp. 145–54.
4. These four non-central functions or other characteristics of the centres are whether the settlement is an incorporated city and whether it possesses a water supply system, a sewage disposal system or an office of an electricity distribution system.
5. 'The Hierarchy of Central Places in Southwestern Wisconsin', *Geogr. Rev.*, vol. 43 (1953) 380–402.
6. R. Vining, 'A Description of Certain Spatial Aspects of an Economic System', *Econ. Development and Cultural Change*, vol. 3 (1955) 167–9.
7. It is a characteristic of the point biserial coefficient of correlation, a statistic designed for the correlation of attributes and variates, that as fewer attributes occur, and as they occur in conjunction with the higher values of the variates, the correlation coefficient increases positively (compare post offices and state liquor stores); conversely, as fewer attributes occur, and in conjunction with the lower values of the variates, the correlation coefficient assures decreasing values (increasing negative values: compare post offices and general stores).
8. These techniques consist of the examination of the reflexive relationships which exist between points on a line and the testing of these for significant differences with the expected random reflexive relationships calculated by Clark and Evans. P. J. Clark and F. C. Evans, 'Distance to Nearest Neighbor as a Measure of Spatial Relations', *Ecology*, vol. 35 (1954) 445–53.

It is interesting to note in relation to the above paper that, as Ambrose[1] *has pointed out, there is a definite break in the hierarchy of town population values. Between the population figures of 974 and 1600, no town exists although a good spread of values appears on either side of this break. In spite of the fact that the authors do not mention this point, it may be conceived as further evidence in support of the notion of a stepped hierarchy in town sizes.*

There are of course other papers which examine and confirm the concept of a hierarchy of central places ranging from the smallest hamlet to the largest metropolitan area. Included in these are studies that consider a hierarchy of functional concentrations, such as shopping centres within large and complex cities. As H. M. Mayer and C. F. Kohn[2] *have pointed out:*

Among the relationships which have been investigated are those of the range of goods and services provided by centres of each given order within a hierarchy, the threshold or minimum population or income required to support a given type of establishment and the relationships of establishments of various types to the sizes of central places within which they occur. Some studies are primarily concerned with the hierarchy of small urban settlements in relation to rural areas which they serve; others, at the opposite end of the scale, are concerned with the role and hinterlands of large metropolises which have regional, continental or world-wide significance. Studies of cities as central places and the development of empirical hypotheses relating size, functions and location of urban settlements constitute a large and important portion of the recent and current work of the urban geographer.

Notes

1. *Analytical Human Geography*, p. 132.
2. *Readings in Urban Geography* (University of Chicago, 1965) p. 184.

SERVICE CENTRES AND THE DISTANCE FACTOR

There is, however, one other aspect of the theme of central-place theory that should be commented upon in more detail in this chapter. This involves not only the concept of hierarchy as examined in the Berry and Garrison paper, but the spatial interpretation of such hierarchies in terms of distance between service centres of like kind.

Christaller dealt with this question of distance both in his basic hexagonal model and in an analysis of empirical results derived from a study of South Germany.

Edward Ullman, in 'A Theory of Location for Cities',[1] *provided a useful summary of Christaller's comments:*

Christaller has recognised typical-size settlements, computed their average population, their distance apart and the size and population of their tributary areas in accordance with his hexagonal theory as Table 22.5 shows. He also states that the number of

TABLE 22.5

	Towns		*Tributary areas*	
Central place	*Distance apart (km.)*	*Population*	*Size (sq. km.)*	*Population*
Market hamlet (*Marktort*)	7	800	45	2,700
Township centre (*Amtsort*)	12	1,500	135	8,100
County seat (*Kreisstadt*)	21	3,500	400	24,000
District city (*Bezirksstadt*)	36	9,000	1,200	75,000
Small state capital (*Gaustadt*)	62	27,000	3,600	225,000
Provincial head city (*Provinzhauptstadt*)	108	90,000	10,800	675,000
Regional capital city (*Landeshauptstadt*)	186	300,000	32,400	2,025,000

central places follows a norm from largest to smallest in the following order: 1:2:6:18:54, etc.[2]

All these figures are computed on the basis of South Germany, but Christaller claims them to be typical for most of Germany and Western Europe. The settlements are classified on the basis of spacing each larger unit in a hexagon of next-order size, so that the distance between similar centres in the table above increases by the $\sqrt{3}$ over the preceding smaller category. The initial distance figure of 7 km. between the smallest centres is chosen because 4–5 km., approximately the distance one can walk in one hour, appears to be a normal service-area limit for the smallest centres. Thus, in a hexagonal scheme, these centres are about 7 km. apart. Cristaller's maps indicate that such centres are spaced close to this norm in South Germany. In the larger categories the norms for distance apart and size of centres appear to be true averages; but

variations from the norm are the rule, although wide discrepancies are not common in the eastern portion of South Germany, which is less highly industrialised than the Rhine–Ruhr areas in the west. The number of central places of each rank varies rather widely from the normal order of expectancy.

The theoretical ideal appears to be most nearly approached in poor, thinly settled farm districts – areas which are most nearly self-contained. In some other sections of Germany industrial concentration seems to be a more important explanation, although elements of the central-place type of distribution are present.

Notes

1. *American Journal of Sociology*, vol. 46 (1941) 138–9.
2. Barnes and Robinson present some interesting maps showing the average distance apart of farmhouses in the driftless area of the Middle West and in southern Ontario. Farmhouses might well be regarded as the smallest settlement units in a central-place scheme, although they might not be in the same numbered sequence. (J. A. Barnes and A. H. Robinson, 'A New Method for the Representation of Dispersed Rural Population', *Geogr. Rev.*, vol. 30 (1940) 134–7.)

Christaller then hints at some of the other factors likely to disturb the theoretical distance interval between service centres of like function in a real landscape. However, for an empirical survey of the spatial hierarchy of central places as related to the distance factor, it is more valuable to turn to a comparatively modern study by J. E. Brush and H. E. Bracey entitled 'Rural Service Centres in Southwestern Wisconsin and Southern England' (1955).[1] *Again the paper is a short one but, unlike that of Berry and Garrison, it presents no problems in terms of understanding. It is quoted in full:*

Comparative analysis of the distribution of rural service centres in south-western Wisconsin and southern England shows that the spatial patterns are alike. Though the two areas are unlike in population density, urbanisation and transportation, and though there are profound differences in settlement history, two orders of service centres exist in both, spaced at about 21-mile and 8- or 10-mile intervals. A third, and still lower, order, spaced at 4- to 6-mile intervals, also appears in both areas. It is impossible to equate the functional importance of rural service centres in Wisconsin and England because of economic and cultural differences. Indeed, distinctive functional types of centres should exist in every major economic or cultural realm on the earth. But the similarities in distribution pattern in Wisconsin and England suggest that there are certain common spatial relationships in the hierarchy of rural service centres.

POPULATION DISTRIBUTION

The nine south-western counties of Wisconsin (Fig. 22.18) and the six southern counties of England (Fig. 22.19) that make up the two areas are situated in regions of moderate relief and fairly uniform rural population distribution. The Wisconsin area, totalling 7170 sq. miles, is a rolling plain or hill land, generally more than a thousand feet above sea level. The surface is formed by dissection of nearly horizontal strata of limestone, shale and sandstone, but has local relief of less than 200 or 300 ft, except for the rugged bluffs near the Mississippi and Wisconsin rivers. Average rural population density is 30 persons to a square mile, ranging from as low as 20 to as high as 40 or 50, especially near the city of Madison. The English area is almost the same size, 6969 sq. miles. There are stretches of highland more than 600 ft above sea level, but the predominant lowlands have local relief of less than 300 ft. The highest parts are in the chalk uplands of Wiltshire, Hampshire and Berkshire, and in the hill country of the Brendons and Exmoor in western Somerset, where abrupt slopes and altitudes of somewhat more than a thousand feet are common. The average rural population density in the six English counties is 182 to a square mile. Villages of several hundred residents and country towns of 2000–15,000 are the rule, in contrast with the single farmsteads and hamlets, villages or small towns of fewer than 5000 in rural Wisconsin.[2] Rural densities as low as 50 to a square mile are found only over considerable areas in the hill country of western Somerset. The total rural population in the six English counties[3] is nearly five times that in the nine Wisconsin counties[4] – 998,000, as compared with 217,000.

The total urban population of the English area, 1,930,000, is more than twelve times that of the Wisconsin area, 158,000. More than two-thirds of the urban residents in the Wisconsin area are in one county, about 100,000 concentrated in Madison and its suburbs. The large cities of the English area, such as Portsmouth, Southampton and Bournemouth on the south cost and Reading

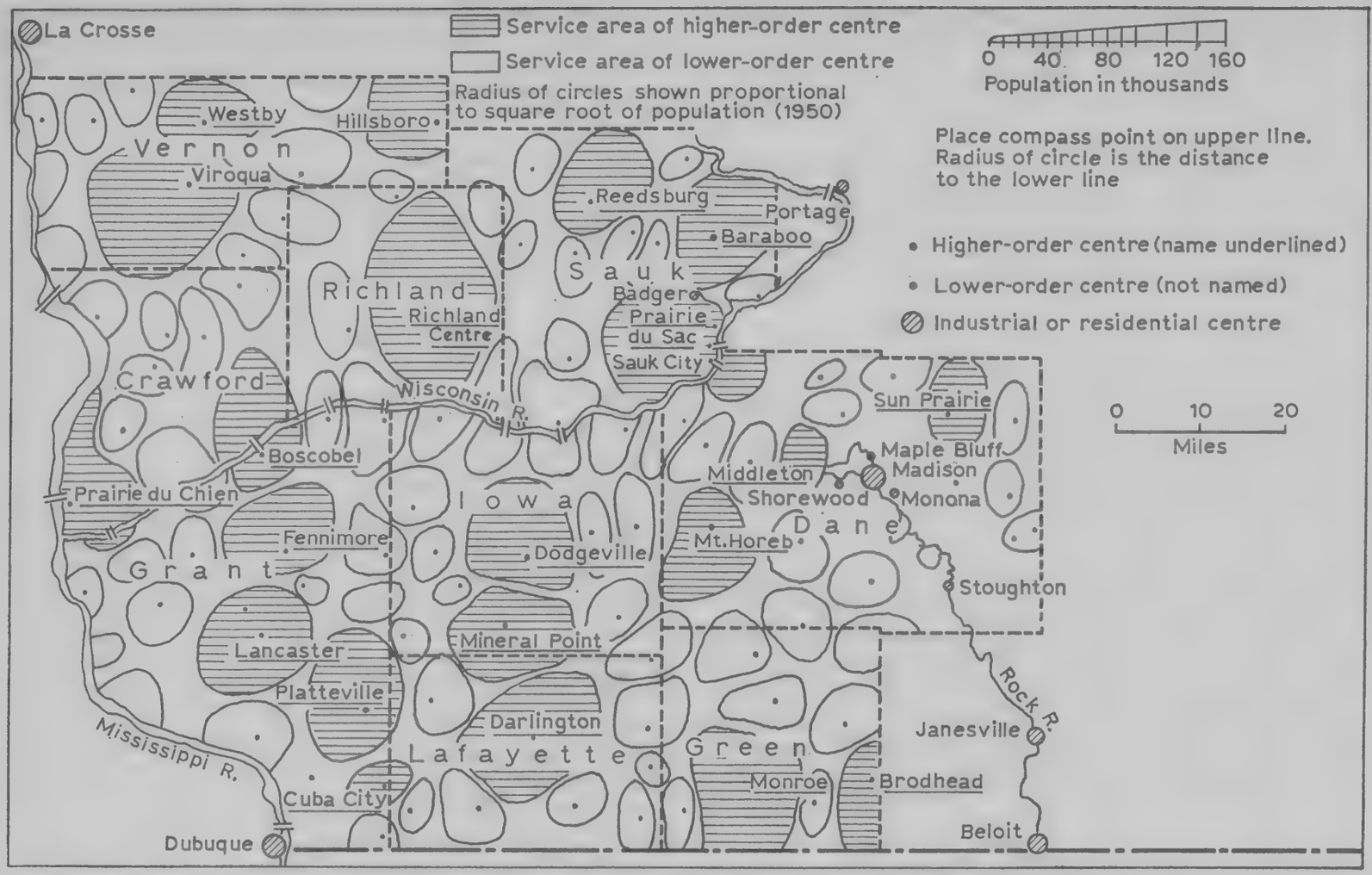


Fig. 22.18 Rural service centres in south-western Wisconsin

and Oxford inland, have each between 100,000 and 300,000 inhabitants and account for about half the urban residents in the six counties. English country towns are generally much larger than Wisconsin towns, but in regard to rural services the dispersion of towns throughout both areas is more significant than the difference in the sizes of rural service centres or the presence of cities in the peripheral areas.

TRANSPORTATION AND ECONOMY

Land occupance is continuous and accessibility good throughout the two areas. Both are covered with farms and are served by close networks of improved highways and secondary roads and by numerous rail lines. The rivers of the English area are small, easily bridged and unnavigable except in their estuaries. In the Wisconsin area, only two rivers form barriers to land transporta-

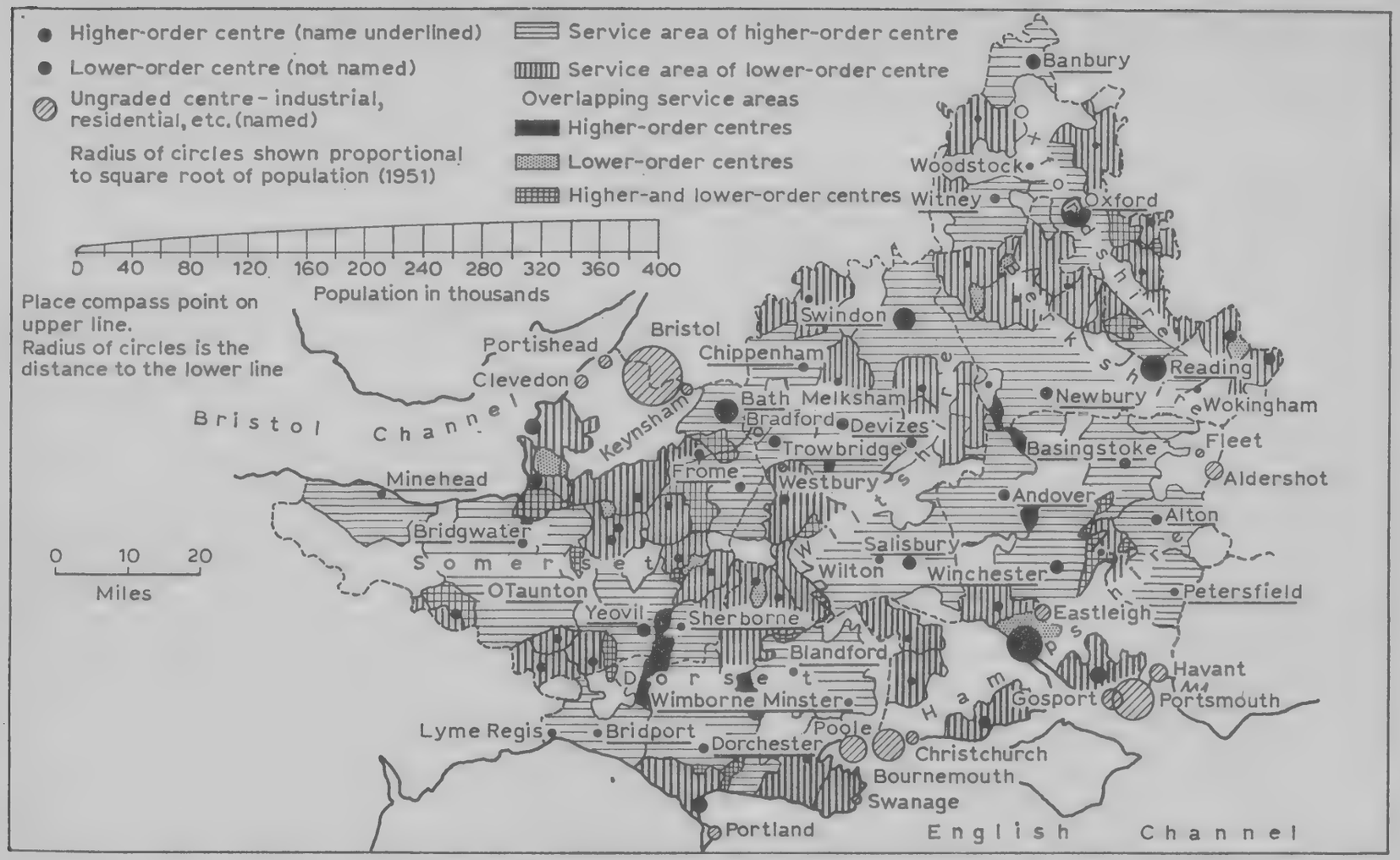


Fig. 22.19 Rural service centres in southern England

tion. The Wisconsin river, no longer used for navigation, is bridged in many places. But the Mississippi, across which few bridges or ferries exist, is a navigable waterway and has had somewhat the same role in transportation as the coasts of the English area. In both areas water transportation is not now important to the development of towns and villages as local trade centres.

The functions of the land transportation system in the two areas differ markedly. In south-western Wisconsin little local use is made of the railways because of competitive services offered by trucking firms and because of the prevalence of private passenger cars and farm trucks. Although the rail lines are neglected and some have been abandoned entirely, the vehicular traffic density on the roads may well be heavier in the Wisconsin area than in the English area. Yet bus services are much better developed in southern England. English towns are served by several scheduled routes on which buses run several times a day. The residents of all except the

most out-of-the-way villages can ordinarily get to one or more towns by bus. In England agricultural produce or freight is carried between farm and local market by road transport, but for longer hauls rail is more often used, though long-distance road haulage is increasing.

In both areas the economic base of the rural population is agriculture, but there are large numbers of rural non-farm people and certain urban-industrial concentrations. South-western Wisconsin farms are predominantly of the dairy type, producing milk for cheese-making and, secondarily, livestock for slaughter. On the heavy clay lands of southern England dairying, with concentration on fresh milk supplies for London and other large cities, is also important. Large acreages, however, are devoted to the cultivation of grain and feedstuffs, stock-raising and fruit-growing.

In the Wisconsin area, Madison with its suburbs forms the only major concentration of manufacturing and other non-agricultural activities. Madison and the minor manufacturing town of Stoughton, in Dane County, mark the westernmost extension of the Rock River manufacturing region. One-half to two-thirds of the working population in the eight other counties are employed directly in agriculture and the remainder in trade, services and professions.

In southern England more persons are employed in agriculture, but they represent a smaller proportion of the total occupied population. As much as 30 per cent of the total working population is employed in manufacturing in Somerset and Wiltshire, as compared with 12 per cent in agriculture. Yet this has not destroyed the close relationships existing between most of the towns and the adjacent rural population. The naval shipyards of Portsmouth, the shipping services and industries of Southampton, and the resort trade of Bournemouth, Clevedon and other coastal places far outweigh their rural service functions. Bristol, with some 440,000 inhabitants, is a still larger urban-industrial concentration, but it lies just outside the limit of the study area and is excluded from consideration, though its suburbs push into Somerset and its influence as a rural service centre is felt over much of the northern part of that county. Reading, Oxford, Bath and Swindon are the only large towns in the area that function as rural service centres and also exhibit significant concentrations of manufacturing or commerce not related to rural service. All towns and villages were included in the survey, but those with slight rural service function were ignored in the subsequent detailed analysis, which embraced the seventy most important rural centres. The six counties are situated well away from the major English manufacturing districts, and the survey area is one of the largest tracts of mixed farming in England. In this respect, therefore, it is comparable with the south-western part of Wisconsin, which lies outside the main manufacturing belt of the United States.

CRITERIA OF CENTRALITY

The role of a place as a trade centre, or its centrality, may be ascertained by either of two methods: (*a*) assessment of the business and services existing in the centre; or (*b*) measurement of the area dependent on the centre for goods and services. In other words, one may ask: Is there a given kind of retail store or service in the centre? Or, alternatively: Where do the people live who go to the centre to obtain a certain commodity or service?

The first method, followed by Dickinson and Smailes in their work in the United Kingdom, is the one applied in south-western Wisconsin.[5] All retail and wholesale trading establishments, banking and financial agencies, trades and personal services, amusements and various other services, including professions and government administration, are considered in the determination of each centre's status. The present-day functional development of service centres in south-western Wisconsin can be assumed to be in accord with the requirements of farmers and other rural people because of the almost complete absence of other urbanising influences. The limits of the area served by a centre are determined by the existence of traffic divides, as shown in the pattern of daily traffic movement towards the centre and towards other centres.

The second method, in which the criterion is use of the centre by rural residents, is the one applied in southern England. This is the method commonly followed by sociological investigators in the United States in their studies of rural community relationships. In the analysis of the six English counties the centres visited by village residents for medical supplies and services, clothing and household goods and banking and entertainment are determined by means of questionnaires. The importance of each service centre is measured by the number of villages partly or wholly dependent on it, and the extent of the service area is taken to include all

villages using the centre for three out of the four broad groups of services. The existence of other facilities and service establishments of the kinds used is implicit in the findings, but no attempt is made to assess them directly.

The advantage of the criterion of use by rural people is not only that it results in a more accurate map of service areas but also that it separates the rural component of centrality from the urban component. This is especially important in assessing the centrality of English cities and towns, some of which have large components of centrality based on urban activities unrelated to the surrounding rural villages. However, development of a majority of the centres in the six English counties – as of those in the nine Wisconsin counties – is closely linked with satisfaction of the needs of rural people.

Working independently, the writers of this paper arranged the centres of their respective survey areas in order of importance. Each identified a group of higher-order centres, though no abrupt break in functions or facilities was recognisable. One out of every five centres in Wisconsin, and one out of three centres in the English area, appeared to belong to a higher order of centrality.

It is clear that significant differences of function exist and that a given type of retail store or specialised service may develop in one country and not in the other. Certain aspects of service function may pertain to higher-order centres in one country and to lower-order centres in the other. A differentiating feature of the higher-order centres in England is their development as shopping centres. More villages depend on the higher-order centres for shopping (retailing of clothing, household goods and various specialised types of consumers' goods) than for professions (medical services and supplies and banking). The second group appears to be fairly well represented in the towns of the lower order. Although Wisconsin towns are also well developed as retail shopping centres, medical and other professional services, except schools, and entertainment are infrequently found in villages. As a result, rural residents in south-western Wisconsin must go to town for specialised commodities except automotive equipment, lumber and hardware and farm equipment, and for all services except banking and secondary schooling.

THE SPATIAL PATTERN OF THE HIERARCHY

In order to demonstrate clearly the existence of higher and lower orders, it is necessary to analyse the distribution of rural centres and their areas. Figs. 22.18 and 22.19 reveal for Wisconsin and England parallel patterns in spacing, service areas and arrangement of higher- and lower-order centres (Table 22.6). Similarities

TABLE 22.6

AVERAGE POPULATION, SERVICE AREAS AND INTERCENTRE DISTANCES, SOUTH-WESTERN WISCONSIN AND SOUTHERN ENGLAND*

	Higher-order centres		*Lower-order centres*	
	19 towns, S.W. Wisconsin	*26 higher dist. centres, S. Engl.*	*73 villages, S.W. Wisconsin*	*44 lower dist. centres, S. Engl.*
Median population†	2,515	13,850	400	5,080
Mean population	3,330	25,950	480	12,425
Mean size of service areas (sq. miles)	129	128	32	48
Mean population of service areas	2,440	21,080	610	7,180
Mean intercentre distance (miles)	21	21	10	8

*The data for south-western Wisconsin apply only to the area west of Madison, excluding eastern Dane County and part of eastern Green County. The service area of a centre is the territory clearly demonstrated by analysis of traffic flow to be dependent for services on that town or village; the ill-defined borders of transition from one service area to another are not included. The service area of a centre in southern England is the territory that includes all villages using the centre for at least three out of four of the following groups of services: shopping, medical supplies and services, business professions and entertainment. There is overlapping of certain adjacent service areas; there is also territory unassigned because the villages are served by very small or very large centres, not belonging in either of the two orders considered here.

†The medians of population provide a better basis for comparison than the means because of the existence of a few large centres in southern England, which unduly weight the means.

in the distribution patterns can be summarised as follows:

1. Higher-order centres occur at a mean distance of 21 miles from one another in both areas.
2. Lower-order centres occur at a mean distance of 10 miles

from one another or from centres of a higher order in Wisconsin and a mean of 8 miles in England.

3. Higher-order centres have service areas of 129 and 128 square miles in Wisconsin and England respectively; lower-order centres have service areas of 32 and 48 square miles respectively.

4. Higher-order centres tend to form clusters or tiers with few or no centres of lower order close to them.

5. Lower-order centres tend to form rows or belts hemmed in by the service areas of higher-order centres and crowded close to one another with their smaller service areas.

Intercentre distances would be similar in both areas if the centres were spaced hexagonally,[6] but the parallels in actual geographic patterns are more significant than this theoretical similarity. The significant features of the spatial patterns which appear in both countries are that the higher-order centres have larger service areas and that the lower-order centres exist only in the regions beyond 8–10 miles from the higher-order centres and not even there if the higher-order centres are as close to one another as 10–12 miles. The map of southern England (Fig. 22.19) shows overlapping of areas served by competing centres. It also shows certain territories not clearly under the influence of centres of either order but served by towns or cities whose rural service functions are too unimportant for them to be ranked with the seventy towns studied. Some of this unassigned territory is served by centres of a third and lower order, which in Wisconsin corresponds to hamlets and in England can be described as service villages. Centres of this third order occur in Somerset at intervals of 4–6 miles from one another or from a centre of a higher order.[7] In Wisconsin they occur at intervals of 5–6 miles. The hamlets in Wisconsin are found chiefly in the transitional territories, left blank on the map (Fig. 22.18), between the well-defined service areas dominated by higher-order centres. The people residing in these transitional territories are dependent on the nearest villages or towns for all except the rudimentary services provided by the hamlets. Service relationships are much more complex in the suburban fringe surrounding Madison and in the lake region extending southwards in Dane County towards Stoughton.

POPULATION AND CENTRALITY

These parallels in spatial hierarchy are clear, though the resident population of centres in both orders is much larger in England (Table 22.6). Towns in south-western Wisconsin have a median population only one-fifth that of the higher-order centres in southern England. The latter show a ratio of about six persons in the centre to five persons in the service area, as compared with seven to five in south-western Wisconsin. Thus it would seem that the larger town population in England correlates with the higher density of rural population. The median population of villages in south-western Wisconsin is less than one-tenth as large as that of their English counterparts. Again, this is in accord with the higher density of population in the service areas in England.

Terms denoting size and rank in the hierarchy in the two countries cannot be used interchangeably. A village of 370 in south-western Wisconsin, such as Barneveld, is a service centre for some 600 persons living on farms within a radius of 4–6 miles. This village is analogous in the hierarchy to a small country town in England, such as Castle Cary, with 2180 inhabitants, which serves the needs of ten or a dozen villages within a radius of 4–6 miles, each with several hundred inhabitants. A well-developed town of about 4600 in south-western Wisconsin, such as Richland Center, serving some 4000 people in the rural territory as far as 10–15 miles around, is a small town by English standards. Although the English centre of Salisbury in Wiltshire, with 32,900 inhabitants, does not serve a larger rural territory, fifty-four villages, with an aggregate rural population of more than 24,000 persons, depend on it for a majority of their ordinary goods and services.

EVOLUTION OF SERVICE CENTRES

The fact that close agricultural settlement has existed in southern England for 1500 years does not seem to have led to any significant difference in the spacing of service centres as they appear today in the two countries, though only 100–125 years have elapsed since the time of agricultural settlement in Wisconsin. Indeed, in certain respects there has been a parallel evolution of centres. During medieval times there were market towns in southern England,

spaced at intervals of 4–6 miles, that served as rudimentary trade centres, accessible by cart roads from the rural villages within an hour's journey.[8] Many English villages are vestigial service centres once of a higher order. In the early nineteenth century, before the coming of railways and automotive vehicles, hamlets also developed in Wisconsin at intervals of 5–6 miles and served as rudimentary trade centres for the farmers living a journey-hour away by wagon roads.

Railway construction in the late nineteenth century had a marked effect in stimulating the growth of trade centres in Wisconsin. The effect of railways was less noticeable in England, but the improvement of roads and the spread of bus services since 1920 appear to have had much the same effect in the growth of large towns as the improvement of roads and the spread of private automobile transportation have had in Wisconsin. Development of professional services, banking and entertainment in Wisconsin has been concentrated in the centres of higher order more than in southern England. This may reflect differences in the relative case of movement in the two countries. Few English workingmen own private cars. Travel by bus increases the journey time and adds to the general inconvenience of a trip to town; this tends to reduce the length and frequency of journey. On the other hand, the persistence of services in lower-order centres may indicate a conservatism on the part of the English – a reluctance to withdraw their patronage from the business people who have served their fathers and grandfathers in satisfying their more personal needs. Some medieval market towns in England continue as service villages just as in Wisconsin the hamlets survive as relict trade centres.

SOME CONCLUSIONS

The fact that analogous geographic patterns have developed in these two widely separated rural territories suggests that the spatial hierarchy of central places is related to distance factors that have a dominant influence in areas of low relief and fairly uniform rural population distribution, despite differences in population density, economic functions and social or political institutions. The fact that rudimentary centres developed at about 6-mile intervals in both areas during the time of primitive transportation leads to the conclusion that the basic distance factor was the time and effort required to get to any trade centre by cart or on foot. The effect of railways was much more significant in Wisconsin than in England because rural settlement and agriculture were still undergoing rapid evolution in the last half of the nineteenth century. The similar effects of automotive vehicles and improved roads in the twentieth century are perhaps unexpected. Although the functional hierarchy is certainly much more variable from place to place and year to year than the rigid system first described by Christaller, it is probable that the spatial relationships observed in Wisconsin and England hold true throughout most of the closely settled parts of north-western Europe. There is need for further comparative analysis of the hierarchy in the countries of eastern and southern Europe and in various parts of the United States in order to determine the limits of variability. There is need also to extend the testing of the principles tentatively presented here to other regions of European settlement such as the British Dominions and Latin America and to regions of non-European settlement in Africa and Asia.

Notes

1. *Geographical Review*, vol. 45, pp. 559–69.

2. The *rural* population in the Wisconsin counties, according to the official definition of *rural* adhered to by the United States Bureau of the Census, includes the inhabitants of small towns and villages where the number is less than 2500 persons as well as the open-country population. In England the meaning of *rural* is not strictly defined by the official census in terms of the number of residents in any closely built-up settlement. Municipal administrative areas such as cities and boroughs are considered urban, though in certain of them the number of inhabitants is only 1000. Civil parishes, some of which comprise more than one village, and some scattered dwellings are considered *rural* if administered by rural district councils, though the aggregate parish population may be as large as 3000 or 4000. One rural parish has as many as 20,000.

3. All population figures cited are taken from *Census* [of] *1951, England and Wales, Preliminary Report* (London: General Register Office, 1951).

4. All population figures cited are taken from U.S. Census of Population, 1950, vol. 1, *Number of Inhabitants* (Washington, D.C.: U.S. Bureau of the Census, 1952) chap. 49, 'Wisconsin'.

5. R. E. Dickinson, 'The Distribution and Functions of the Smaller Settlements of East Anglia', *Geography*, vol. 17 (1932) 19–31.
A. E. Smailes, 'The Urban Hierarchy of England and Wales' *Geography*, vol. 29 (1944) 41–51.

6. If the seventy rural service centres in the six English counties were spaced at equal distances in a perfect hexagonal pattern, they would be

10·7 miles apart. In the part of south-western Wisconsin west of Madison, the ninety-two centres of village or town rank would be 10·0 miles apart if spaced in a perfect hexagonal pattern. The formula used in these computations is as follows:

$$D = 1{\cdot}07\sqrt{\frac{A}{n}}$$

where D is the distance between the centres, A is the total area within which the centres are dispersed and n is the number of centres. The formula calls for nothing more than the square root of the density of centres within the area multiplied by a constant factor and should not be considered as giving a true measurement of scattering or dispersion.

7. Dickinson found there are 'small towns' with 1000–2000 population in East Anglia, spaced at about 4-mile intervals detail on [t.s.].

8. A journey-hour on foot or by cart, travelling at the rate of 3 miles an hour. R. E. Dickinson, *City, Region and Regionalism* (London, 1947).

CENTRAL-PLACE THEORY EVALUATED

In conclusion, it cannot but be agreed that in spite of the various reservations and criticisms to which attention has been drawn, central-place theory remains an important expression of those basic principles which account for settlement patterns.

Ullman[1] *has written:*

Even in the closely articulated national economy of the United States there are strong forces at work to produce a central-place distribution of settlements. It is true that products under our national economy are characteristically shipped from producing areas through local shipping points directly to consuming centres which are often remote. However, the distribution of goods or imports brought into an area is characteristically carried on through brokerage, wholesale, and retail channels in central cities. This graduated division of functions supports a central-place framework of settlements. Many non-industrial regions of relatively uniform land surface have cities distributed so evenly over the land that some sort of central-place theory appears to be the prime explanation.[2] It should be worth while to study this distribution and compare it with other areas.[3] In New England, on the other hand, where cities are primarily industrial centres based on distant raw materials and extraregional markets, instead of the land's supporting the city the reverse is more nearly true: the city supports the countryside by providing a market for farm products, and thus infertile rural areas are kept from being even more deserted than they are now.

The forces making for concentration at certain places and the inevitable rise of cities at these favoured places have been emphasised by geographers and other scholars. The phenomenal growth of industry and world trade in the last hundred years and the concomitant growth of cities justify this emphasis but have perhaps unintentionally caused the intimate connection between a city and its surrounding area partially to be overlooked. Explanation in terms of concentration is most important for industrial districts but does not provide a complete areal theory for distribution of settlements. Furthermore, there is evidence that 'of late . . . the rapid growth of the larger cities has reflected their increasing importance as commercial and service centres rather than as industrial centres'.[4] Some form of the central-place theory should provide the most realistic key to the distribution of settlements where there is no marked concentration – in agricultural areas where explanation has been most difficult in the past. For all areas the system may well furnish a theoretical norm from which deviations may be measured.[5] It might also be an aid in planning the development of new areas. If the theory is kept in mind by workers in academic and planning fields as more studies are made, its validity may be tested and its structure refined in accordance with regional differences.

Notes

1. Op. cit., p. 208.

2. For a confirmation of this, see the column diagram on p. 73 of Lösch, op. cit., which shows the minimum distances between towns in Iowa of three different size classes. The maps of trade-centre distribution in the works of Zimmerman, Smith and Landis also show an even spacing of centres.

3. The following table gives the average community area for 140 villages in the United States in 1930. In the table, notice throughout that (1) the larger the village, the larger its tributary area in each region, and (2) the sparser the rural population density, the larger the village tributary area for each size class (contrast mid-Atlantic with Far West, etc.).

	Community area in square miles		
Region	*Small villages (250–1000 pop.)*	*Medium villages (1000–1750 pop.)*	*Large villages (1750–2500 pop.)*
Mid-Atlantic	43	46	87
South	77	111	146
Middle West	81	113	148
Far West	..	365	223

Although 140 is only a sample of the number of villages in the country, the figures are significant because the service areas were carefully and uniformly delimited in the field for all villages.

In New York 26 square miles was found to be the average area per village in 1920. 'Village' refers to any settlement under 2500 population. Nearness to cities, type of agriculture and routes of travel are cited as the three most important factors influencing density of villages. Since areas near cities are suburbanised in some cases, as around New York City, the village-density in these districts is correspondingly high. Some urban counties with smaller cities (Rochester, Syracuse and Niagara Falls) have few suburbs, and consequently the villages are farther apart than in many agricultural counties.

In sample areas of New York State the average distance from a village of 250 or under to another of the same size or larger is about 3 miles; for the 250–749 class it is 3–5 miles; for the 750–1249 class, 5–7 miles.

Kolb makes some interesting suggestions as to the distances between centres. He shows that spacing is closer in central Wisconsin than in Kansas, which is more sparsely settled. J. H. Kolb, 'Service Relations of Town and Country', *Wisconsin Agricultural Experimental Station Res. Bull.*, vol. 58 (1923).

In Iowa, 'the dominant factor determining the *size* of convenience-goods areas is distance'. (Second State of Iowa Planning Board Report, April 1935, p. 198); this report contains fertile suggestions of trade areas for Iowa towns. Valuable detailed reports on retail trade areas for some Iowa counties have also been made by the same agency.

4. U.S. National Resources Committee, 'Our Cities – their Role in the National Economy', *Report of the Urbanism Committee* (Washington, 1937).

5. Some form of the central-place concept might well be used to advantage in interpreting the distribution of outlying business districts in cities.

As an interesting corollary to Ullman's closing words, much more recently Ambrose[1] *was able to say of central-place theory that:*

Clearly we cannot expect it to produce a complete explanation of the present size and spacing of central places, since so many important explanatory factors are necessarily omitted. It is, however, a theory that contains certain vitally important concepts, such as the interdependence of town and region, the hierarchy of functions and centres, threshold population and spatial competition. A great deal of research has been built on these central ideas and in at least one region, the newly reclaimed Dutch polders, new service centres have been located in accordance with the principles embodied in the theory.

Note

1. *Analytical Human Geography*, p. 129.

There is little doubt among geographers as to the value of Christaller's work. William Bunge[1] *perhaps rates it particularly highly:*

If it were not for the existence of central-place theory it would not be possible to be so emphatic about the existence of a theoretical geography independent of any set of mother sciences. Geography is a basic science since it produces new theory and the proof of that assertion lies most clearly in the existence of central-place theory. With the possible exception of cartography, this author is of the opinion that the initial and growing beauty of central-place theory is geography's finest intellectual product and puts Christaller in a place of great honour.

Note

1. *Theoretical Geography* (1966) p. 133.

To introduce the student to political geography we have selected two papers so diverse that at first sight they may seem almost incompatible, and our choice mischievously perverse. Mackinder's paper, published as long ago as 1904, essays 'a correlation between the larger geographical and the larger historical generalisations'. The interesting oral discussion at the Royal Geographical Society is also reproduced here. After World War I his 'Democratic Ideals and Reality' (1919) elaborated these ideas, referring to a great core area of Eurafrasia as the Heartland, and suggesting that:

Who rules East Europe commands the Heartland.
Who rules the Heartland commands the World-Island.
Who rules the World-Island commands the World.

These ideas seem to have affected strategic thinking in both Russia and Germany, including German notions on the alleged organic nature of the State and its claim for 'lebensraum', notably by Karl Haushofer (1869–1946), merging in turn with Nazi doctrine on 'Geopolitik', the need for a German-dominated World State. Any standard work on political geography will take the interested reader farther into the continuing debate on Mackinder's ideas into the age of long-range aircraft, rocketry and nuclear weapons. Here we include a short extract from H. de Blij's 'Systematic Political Geography' on 'The Heartland Debate'.

House's paper, presented orally sixty years later and published in 1968, is a product of quite a different scale of operation, and an avowedly different approach. Culminating in two case studies of frontier areas in Europe including international boundaries – contrasted but both subject to change – House pleads for a patient building of political geography upon just such detailed studies, as against 'those premature categorisations beloved by, and indeed more appropriate to the military school of boundary defenders and delineators'.

THE GEOGRAPHICAL PIVOT OF HISTORY[1]

H. J. Mackinder

When historians in the remote future come to look back on the group of centuries through which we are now passing, and see them foreshortened, as we today see the Egyptian dynasties, it may well be that they will describe the last 400 years as the Columbian epoch, and will say that it ended soon after the year 1900. Of late it has been a commonplace to speak of geographical exploration as nearly over, and it is recognised that geography must be diverted to the purpose of intensive survey and philosophic synthesis. In 400 years the outline of the map of the world has been completed with approximate accuracy, and even in the polar regions the voyages of Nansen and Scott have very narrowly reduced the last possibility of dramatic discoveries. But the opening of the twentieth century is appropriate as the end of a great historic epoch, not merely on account of this achievement, great though it be. The missionary, the conqueror, the farmer, the miner and, of late, the engineer, have followed so closely in the traveller's footsteps that the world, in its remoter borders, has hardly been revealed before we must chronicle its virtually complete political appropriation. In Europe, North America, South America, Africa and Australasia there is scarcely a region left for the pegging out of a claim of ownership, unless as the result of a war between civilised or half-civilised powers. Even in Asia we are probably witnessing the last moves of the game first played by the horsemen of Yermak the Cossack and the shipmen of Vasco da Gama. Broadly speaking, we may contrast the Columbian epoch with the age which preceded it, by describing its essential characteristic as the expansion of Europe against almost negligible resistances, whereas medieval Christendom was pent into a narrow region and threatened by external barbarism. From the present time forth, in the post-Columbian age, we shall again have to deal with a closed political system, and none the less that it will be one of world-wide scope. Every explosion of social forces, instead of being dissipated in a surrounding circuit of unknown space and barbaric chaos, will be sharply re-echoed from the far side of the globe, and weak elements in the political and economic organisms of the world will be shattered in consequence. There is a vast difference of effect in the fall of a shell into an earthwork and its fall amid the closed spaces and rigid structures of a great building or ship.

Probably some half-consciousness of this fact is at last diverting much of the attention of statesmen in all parts of the world from territorial expansion to the struggle for relative efficiency.

It appears to me, therefore, that in the present decade we are for the first time in a position to attempt, with some degree of completeness, a correlation between the larger geographical and the larger historical generalisations. For the first time we can perceive something of the real proportion of features and events on the stage of the whole world, and may seek a formula which shall express certain aspects, at any rate, of geographical causation in universal history. If we are fortunate, that formula should have a practical value as setting into perspective some of the competing forces in current international politics. The familiar phrase about the westward march of empire is an empirical and fragmentary attempt of the kind. I propose this evening describing those physical features of the world which I believe to have been most coercive of human action, and presenting some of the chief phrases as organically connected with them, even in the ages when they were unknown to geography. My aim will not be to discuss the influence of this or that kind of feature, or yet to make a study in regional geography, but rather to exhibit human history as part of the life of the world organism. I recognise that I can only arrive at one aspect of the truth, and I have no wish to stray into excessive materialism. Man and not nature initiates, but nature in large measure controls. My concern is with the general physical control, rather than the causes of universal history. It is obvious that only a first approximation to truth can be hoped for. I shall be humble to my critics.

The late Prof. Freeman held that the only history which counts is that of the Mediterranean and European races. In a sense, of course, this is true, for it is among these races that have originated the ideas which have rendered the inheritors of Greece and Rome dominant throughout the world. In another and very important sense, however, such a limitation has a cramping effect upon thought. The ideas which go to form a nation, as opposed to a mere crowd of human animals, have usually been accepted under the pressure of a common tribulation, and under a common necessity of resistance to external force. The idea of England was beaten into the Heptarchy by Danish and Norman conquerors; the idea of France was forced upon competing Franks, Goths, and Romans by the Huns at Châlons, and in the Hundred Years' War with England; the idea of Christendom was born of the Roman persecutions, and matured by the Crusades; the idea of the United States was accepted, and local colonial patriotism sunk, only in the long War of Independence; the idea of the German Empire was reluctantly adopted in South Germany only after a struggle against France in comradeship with North Germany. What I may describe as the literary conception of history, by concentrating attention upon ideas and upon the civilisation which is their outcome, is apt to lose sight of the more elemental movements whose pressure is commonly the exciting cause of the efforts in which great ideas are nourished. A repellent personality performs a valuable social function in uniting his enemies, and it was under the pressure of external barbarism that Europe achieved her civilisation. I ask you, therefore, for a moment to look upon Europe and European history as subordinate to Asia and Asiatic history, for European civilisation is, in a very real sense, the outcome of the secular struggle against Asiatic invasion.

The most remarkable contrast in the political map of modern Europe is that presented by the vast area of Russia occupying half the Continent and the group of smaller territories tenanted by the Western Powers. From a physical point of view, there is, of course, a like contrast between the unbroken lowland of the east and the rich complex of mountains and valleys, islands and peninsulas, which together form the remainder of this part of the world. At first sight it would appear that in these familiar facts we have a correlation between natural environment and political organisation so obvious as hardly to be worthy of description, especially when we note that throughout the Russian plain a cold winter is opposed to a hot summer, and the conditions of human existence thus rendered additionally uniform. Yet a series of historical maps, such as that contained in the *Oxford Atlas*, will reveal the fact that not merely is the rough coincidence of European Russia with the Eastern Plain of Europe a matter of the last hundred years or so, but that in all earlier time there was persistent re-assertion of quite another tendency in the political grouping. Two groups of states usually divided the country into northern and southern political systems. The fact is that the orographical map does not express the particular physical contrast which has until very lately controlled human movement and

settlement in Russia. When the screen of winter snow fades northward off the vast face of the plain, it is followed by rains whose maximum occurs in May and June beside the Black Sea, but near the Baltic and White Seas is deferred to July and August. In the

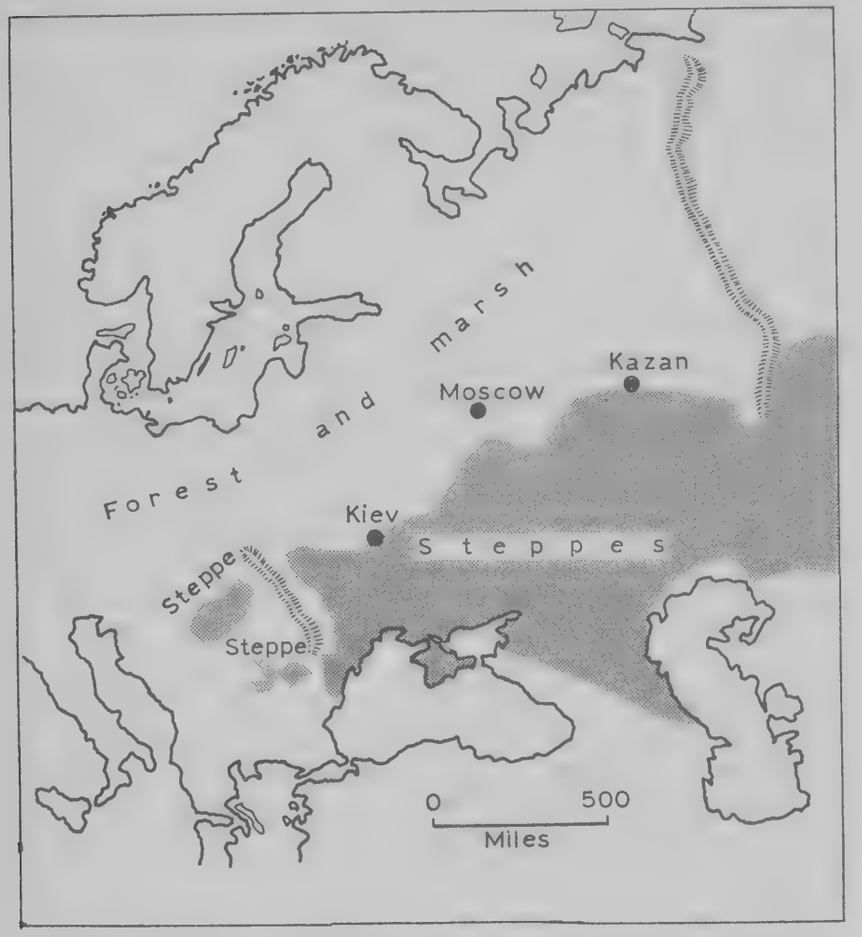


Fig. 23.1 Eastern Europe before the nineteenth century (after Drude in *Berghaus's Physical Atlas*)

south the later summer is a period of drought. As a consequence of this climatic regime, the north and north-west were forest broken only by marshes, whereas the south and south-east were a boundless grassy steppe, with trees only along the rivers. The line separating the two regions ran diagonally north-eastward from the northern end of the Carpathians to a point in the Ural range nearer to its southern than to its northern extremity. Moscow lies a little to the north of this line, or, in other words, on the forest side of it. Outside Russia the boundary of the great forest ran

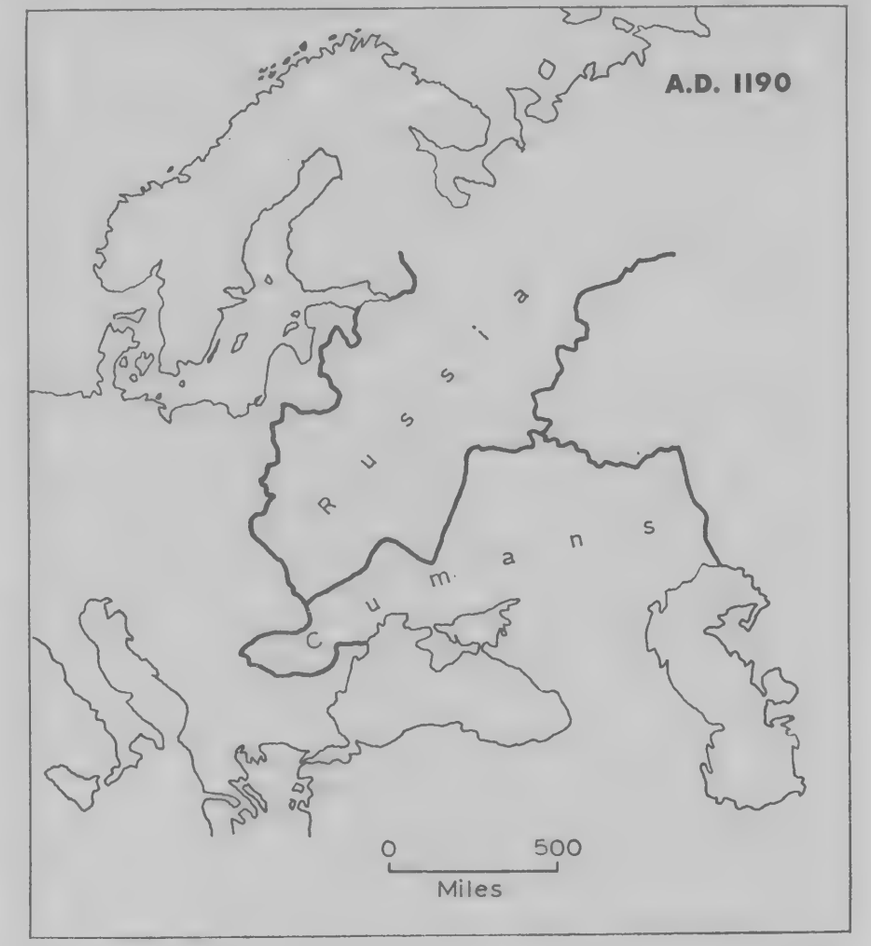


Fig. 23.2 Political divisions of Eastern Europe at the time of the Third Crusade (after the *Oxford Historical Atlas*)

westward almost exactly through the centre of the European isthmus, which is 800 miles across between the Baltic and the Black Seas. Beyond this, in Peninsular Europe, the woods spread on through the plains of Germany in the north, while the steppe lands in the south turned the great Transylvanian bastion of the Car-

pathians, and extended up the Danube, through what are now the cornfields of Roumania, to the Iron Gates. A detached area of steppes, known locally as Pusztas, now largely cultivated, occupied the plain of Hungary, ingirt by the forested rim of Carpathian and Alpine mountains. In all the west of Russia, save in the far north, the clearing of the forests, the drainage of the marshes and the tillage of the steppes have recently averaged the character of the landscape, and in large measure obliterated a distinction which was formerly very coercive of humanity.

The earlier Russia and Poland were established wholly in the glades of the forest. Through the steppe on the other hand there came from the unknown recesses of Asia, by the gateway between the Ural Mountains and the Caspian Sea, in all the centuries from the fifth to the sixteenth, a remarkable succession of Turanian nomadic peoples – Huns, Avars, Bulgarians, Magyars, Khazars, Patzinaks, Cumans, Mongols, Kalmuks. Under Attila the Huns established themselves in the midst of the Pusztas, in the uttermost Danubian outlier of the steppes, and thence dealt blows northward, westward and southward against the settled peoples of Europe. A large part of modern history might be written as a commentary upon the changes directly or indirectly ensuing from these raids. The Angles and Saxons, it is quite possible, were then driven to cross the seas to found England in Britain. The Franks, the Goths and the Roman provincials were compelled, for the first time, to stand shoulder to shoulder on the battlefield of Châlons, making common cause against the Asiatics, who were unconsciously welding together modern France. Venice was founded from the destruction of Aquileia and Padua; and even the Papacy owed a decisive prestige to the successful mediation of Pope Leo with Attila at Milan. Such was the harvest of results produced by a cloud of ruthless and idealess horsemen sweeping over the unimpeded plain – a blow, as it were, from the great Asiatic hammer striking freely through the vacant space. The Huns were followed by the Avars. It was for a marchland against these that Austria was founded, and Vienna fortified, as the result of the campaigns of Charlemagne. The Magyar came next, and by incessant raiding from his steppe base in Hungary increased the significance of the Austrian outpost, so drawing the political focus of Germany eastward to the margin of the realm. The Bulgarian established a ruling caste south of the Danube, and has left his name upon the map, although his language has yielded to that of his Slavonic subjects. Perhaps the longest and most effective occupation of the Russian steppe proper was that of the Khazars, who were contemporaries of the great Saracen movement: the

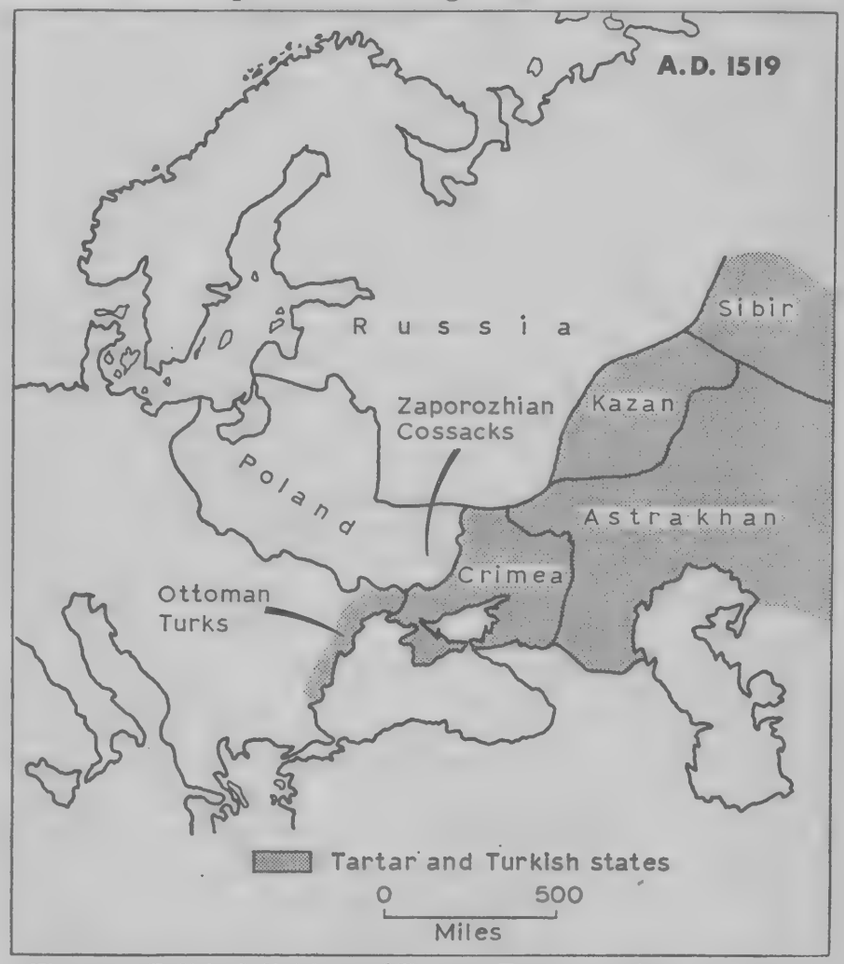


Fig. 23.3 Political divisions of Eastern Europe at the accession of Charles V (after the *Oxford Historical Atlas*)

Arab geographers knew the Caspian as the Khazar Sea. In the end, however, new hordes arrived from Mongolia, and for two centuries Russia in the northern forest was held tributary to the Mongol Khans of Kipchak, or 'the Steppe', and Russian development was thus delayed and biased at a time when the remainder of Europe was rapidly advancing.

It should be noted that the rivers running from the Forest to the Black and Caspian Seas cross the whole breadth of the steppe-land path of the nomads, and that from time to time there were transient movements along their courses at right-angles to the movement of the horsemen. Thus the missionaries of Greek Christianity ascended the Dnieper to Kiev, just as beforehand the Norse Varangians had descended the same river on their way to Constantinople. Still earlier, the Teutonic Goths appear for a moment upon the Dniester, having crossed Europe from the shores of the Baltic in the same south-eastward direction. But these are passing episodes which do not invalidate the broader generalisation. For a thousand years a series of horse-riding peoples emerged from Asia through the broad interval between the Ural Mountains and the Caspian Sea, rode through the open spaces of southern Russia and struck home into Hungary in the very heart of the European peninsula, shaping by the necessity of opposing them the history of each of the great peoples around – the Russians, the Germans, the French, the Italians and the Byzantine Greeks. That they stimulated healthy and powerful reaction, instead of crushing opposition under a widespread despotism, was due to the fact that the mobility of their power was conditioned by the steppes, and necessarily ceased in the surrounding forests and mountains.

A rival mobility of power was that of the Vikings in their boats. Descending from Scandinavia both upon the northern and the southern shores of Europe, they penetrated inland by the river ways. But the scope of their action was limited, for, broadly speaking, their power was effective only in the neighbourhood of the water. Thus the settled peoples of Europe lay gripped between two pressures – that of the Asiatic nomads from the east, and on the other three sides that of the pirates from the sea. From its very nature neither pressure was overwhelming, and both therefore were stimulative. It is noteworthy that the formative influence of the Scandinavians was second only in significance to that of the nomads, for under their attack both England and France made long moves towards unity, while the unity of Italy was broken by them. In earlier times, Rome had mobilised the power of her settled peoples by means of her roads, but the Roman roads had fallen into decay, and were not replaced until the eighteenth century.

It is likely that even the Hunnish invasion was by no means the first of the Asiatic series. The Scythians of the Homeric and Herodotian accounts, drinking the milk of mares, obviously practised the same arts of life, and were probably of the same race as the later inhabitants of the steppe. The Celtic element in the river-names *Don*, *Don*etz, *Dn*ieper, *Dn*iester and *Dan*ube may possibly betoken the passage of peoples of similar habits, though not of identical race, but it is not unlikely that the Celts came merely from the northern forests, like the Goths and Varangians of a later time. The great wedge of population, however, which the anthropologists characterise as brachycephalic, driven westward from brachycephalic Asia through Central Europe into France, is apparently intrusive between the northern, western and southern dolicocephalic populations, and may very probably have been derived from Asia.

The full meaning of Asiatic influence upon Europe is not, however, discernible until we come to the Mongol invasions of the fifteenth century; but before we analyse the essential facts concerning these, it is desirable to shift our geographical viewpoint from Europe, so that we may consider the Old World in its entirety. It is obvious that, since the rainfall is derived from the sea, the heart of the greatest land-mass is likely to be relatively dry. We are not, therefore, surprised to find that two-thirds of all the world's population is concentrated in relatively small areas along the margins of the great continent – in Europe, beside the Atlantic Ocean; in the Indies and China, beside the Indian and Pacific Oceans. A vast belt of almost uninhabited, because practically rainless, land extends as the Sahara completely across northern Africa into Arabia. Central and southern Africa were almost as completely severed from Europe and Asia throughout the greater part of history as were the Americas and Australia. In fact, the southern boundary of Europe was and is the Sahara rather than the Mediterranean, for it is the desert which divides the black man from the white. The continuous land-mass of Euro-Asia thus included between the ocean and the desert measures 21 million square miles, or half of all the land on the globe, if we exclude from reckoning the deserts of Sahara and Arabia. There are many detached deserts scattered through Asia, from Syria and Persia north-eastward to Manchuria, but no such continuous vacancy as to be comparable with the Sahara. On the other hand, Euro-

Asia is characterised by a very remarkable distribution of river drainage. Throughout an immense portion of the centre and north, the rivers have been practically useless for purposes of human communication with the outer world. The Volga, the Oxus and the Jaxartes drain into salt lakes; the Ob, the Yenisei and the Lena into the frozen ocean of the north. These are six of the greatest rivers in the world. There are many smaller but still considerable streams in the same area, such as the Tarim and the Helmand, which similarly fail to reach the ocean. Thus the core

Fig. 23.4 Continental and arctic drainage (equal-area projection)

of Euro-Asia, although mottled with desert patches, is on the whole a steppe-land supplying a widespread if often scanty pasture, and there are not a few river-fed oases in it, but it is wholly unpenetrated by waterways from the ocean. In other words, we have in this immense area all the conditions for the maintenance of a sparse, but in the aggregate considerable, population of horse-riding and camel-riding nomads. Their realm is limited northward by a broad belt of subarctic forest and marsh, wherein the climate is too rigorous, except at the eastern and western extremities, for the development of agricultural settlements. In the east the forests extend southward to the Pacific coast in the Amur land and Manchuria. Similarly in the west, in prehistoric Europe, forest was the predominant vegetation. Thus framed in to the north-east, north and north-west, the steppes spread continuously for 4000 miles from the Pusztas of Hungary to the Little Gobi of Manchuria, and, except in their westernmost extremity, they are untraversed by rivers draining to an accessible ocean, for we may neglect the very recent efforts to trade to the mouths of the Ob and Yenisei. In Europe, western Siberia and western Turkestan the steppe lands lie low, in some places below the level of the sea. Farther to the east, in Mongolia, they extend over plateaux; but the passage from the one level to the other, over the naked, unscarped lower ranges of the arid heartland, presents little difficulty.

The hordes which ultimately fell upon Europe in the middle of the fourteenth century gathered their first force 3000 miles away on the high steppes of Mongolia. The havoc wrought for a few years in Poland, Silesia, Moravia, Hungary, Croatia and Serbia was, however, but the remotest and the most transient result of the great stirring of the nomads of the East associated with the name of Genghis Khan. While the Golden Horde occupied the steppe of Kipchak, from the Sea of Aral, through the interval between the Ural range and the Caspian, to the foot of the Carpathians, another horde, descending south-westward between the Caspian Sea and the Hindu Kush into Persia, Mesopotamia and even into Syria, founded the domain of the Ilkhan. A third subsequently struck into northern China, conquering Cathay. India and Mangi, or southern China, were for a time sheltered by the incomparable barrier of Tibet, to whose efficacy there is, perhaps, nothing similar in the world, unless it be the Sahara desert and the polar ice. But at a later time, in the days of Marco Polo in the case of Mangi, in those of Tamerlane in the case of India, the obstacle was circumvented. Thus it happened that in this typical and well-recorded instance, all the settled margins of the Old World sooner or later felt the expansive force of mobile power originating in the steppe. Russia, Persia, India and China were either made tributary, or received Mongol dynasties. Even the incipient power of the Turks in Asia Minor was struck down for half a century.

As in the case of Europe, so in other marginal lands of Euro-Asia there are records of earlier invasions. China had more than once to submit to conquest from the north; India several times to conquest from the north-west. In the case of Persia, however, at least one of the earlier descents has a special significance in the history of Western civilisation. Three or four centuries before the

Mongols, the Seljuk Turks, emerging from Central Asia, overran by this path an immense area of the land, which we may describe as of the five seas – Caspian, Black, Mediterranean, Red and Persian. They established themselves at Kerman, at Hamadan and in Asia Minor, and they overthrew the Saracen dominion of Baghdad and Damascus. It was ostensibly to punish their treatment of the Christian pilgrims at Jerusalem that Christendom undertook the great series of campaigns known collectively as the Crusades. Although these failed in their immediate objects, they so stirred and united Europe that we may count them as the beginning of modern history – another striking instance of European advance stimulated by the necessity of reacting against pressure from the heart of Asia.

The conception of Euro-Asia to which we thus attain is that of a continuous land, ice-girt in the north, water-girt elsewhere, measuring 21 million square miles, or more than three times the area of North America, whose centre and north, measuring some 9 million square miles, or more than twice the area of Europe, have no available waterways to the ocean, but, on the other hand, except in the subarctic forest, are very generally favourable to the mobility of horsemen and camelmen. To east, south and west of this heartland are marginal regions, ranged in a vast crescent, accessible to shipmen. According to physical conformation, these regions are four in number, and it is not a little remarkable that in a general way they respectively coincide with the spheres of the four great religions – Buddhism, Brahminism, Mahometanism and Christianity. The first two are the monsoon lands, turned the one towards the Pacific, and the other towards the Indian Ocean. The fourth is Europe, watered by the Atlantic rains from the west. These three together, measuring less than 7 million square miles, have more than 1000 million people, or two-thirds of the world population. The third, coinciding with the land of the Five Seas, or, as it is more often described, the Nearer East, is in large measure deprived of moisture by the proximity of Africa, and, except in the oases, is therefore thinly peopled. In some degree it partakes of the characteristics both of the marginal belt and of the central area of Euro-Asia. It is mainly devoid of forest, is patched with desert, and is therefore suitable for the operations of the nomad. Dominantly, however, it is marginal, for sea-gulfs and oceanic rivers lay it open to sea-power, and permit of the exercise of such power from it. As a consequence, periodically throughout history, we have here had empires belonging essentially to the marginal series, based on the agricultural populations of the great oases of Babylonia and Egypt, and in free water communication with the civilised worlds of the Mediterranean and the Indies. But, as we should expect, these empires have been subject to an unparalleled series of revolutions, some due to Scythian, Turkish and Mongol raids from Central Asia, others to the effort of the Mediterranean peoples to conquer the overland ways from the western to the eastern ocean. Here is the weakest spot in the girdle of early civilisations, for the isthmus of Suez divided seapower into eastern and western, and the arid wastes of Persia advancing from Central Asia to the Persian Gulf gave constant opportunity for nomad-power to strike home to the ocean edge, dividing India and China, on the one hand, from the Mediterranean world on the other. Whenever the Babylonian, the Syrian and the Egyptian oases were weakly held, the steppe peoples could treat the open tablelands of Iran and Asia Minor as forward posts whence to strike through the Punjab into India, through Syria into Egypt and over the broken bridge of the Bosphorus and Dardanelles into Hungary. Vienna stood in the gateway of Inner Europe, withstanding the nomadic raids, both those which came by the direct road through the Russian steppe, and those which came by the loop way to the south of the Black and Caspian Seas.

Here we have illustrated the essential difference between the Saracen and the Turkish controls of the Nearer East. The Saracens were a branch of the Semitic race, essentially peoples of the Euphrates and Nile and of the smaller oases of Lower Asia. They created a great empire by availing themselves of the two mobilities permitted by their land – that of the horse and camel on the one hand, that of the ship on the other. At different times their fleets controlled both the Mediterranean as far as Spain, and the Indian Ocean to the Malay islands. From their strategically central position between the eastern and western oceans, they attempted the conquest of all the marginal lands of the Old World, imitating Alexander and anticipating Napoleon. They could even threaten the steppe land. Wholly distinct from Arabia as from Europe, India and China were the Turanian pagans from the closed heart of Asia, the Turks who destroyed the Saracen civilisation.

Mobility upon the ocean is the natural rival of horse and camel

mobility in the heart of the continent. It was upon navigation of oceanic rivers that was based the Potamic stage of civilisation, that of China on the Yangtse, that of India on the Ganges, that of Babylonia on the Euphrates, that of Egypt on the Nile. It was essentially upon the navigation of the Mediterranean that was based what has been described as the Thalassic stage of civilisation, that of the Greeks and Romans. The Saracens and the Vikings held sway by navigation of the oceanic coasts.

The all-important result of the discovery of the Cape road to the Indies was to connect the western and eastern coastal navigations of Euro-Asia, even though by a circuitous route, and thus in some measure to neutralise the strategical advantage of the central position of the steppe nomads by pressing upon them in the rear. The revolution commenced by the great mariners of the Columbian generation endowed Christendom with the widest possible mobility of power, short of a winged mobility. The one and continuous ocean enveloping the divided and insular lands is, of course, the geographical condition of ultimate unity in the command of the sea, and of the whole theory of modern naval strategy and policy as expounded by such writers as Captain Mahan and Mr Spenser Wilkinson. The broad political effect was to reverse the relations of Europe and Asia, for whereas in the Middle Ages Europe was caged between an impassable desert to the south, an unknown ocean to the west and icy or forested wastes to the north and north-east, and in the east and south-east was constantly threatened by the superior mobility of the horsemen and camelmen, she now emerged upon the world, multiplying more than thirty-fold the sea surface and coastal lands to which she had access, and wrapping her influence round the Euro-Asiatic landpower which had hitherto threatened her very existence. New Europes were created in the vacant lands discovered in the midst of the waters, and what Britain and Scandinavia were to Europe in the earlier time, that have America and Australia, and in some measure even trans-Saharan Africa, now become to Euro-Asia. Britain, Canada, the United States, South Africa, Australia and Japan are now a ring of outer and insular bases for sea-power and commerce, inaccessible to the land-power of Euro-Asia.

But the land-power still remains, and recent events have again increased its significance. While the maritime peoples of Western Europe have covered the ocean with their fleets, settled the outer continents and in varying degree made tributary the oceanic margins of Asia, Russia has organised the Cossacks, and, emerging from her northern forests, has policed the steppe by setting her own nomads to meet the Tartar nomads. The Tudor century, which saw the expansion of Western Europe over the sea, also saw Russian power carried from Moscow through Siberia. The eastward swoop of the horsemen across Asia was an event almost as pregnant with political consequences as was the rounding of the Cape, although the two movements long remained apart.

It is probably one of the most striking coincidences of history that the seaward and the landward expansion of Europe should, in a sense, continue the ancient opposition between Roman and Greek. Few great failures have had more far-reaching consequences than the failure of Rome to Latinise the Greek. The Teuton was civilised and Christianised by the Roman, the Slav in the main by the Greek. It is the Romano-Teuton who in later times embarked upon the ocean; it was the Graeco-Slav who rode over the steppes, conquering the Turanian. Thus the modern land-power differs from the sea-power no less in the source of its ideals than in the material conditions of its mobility.[2]

In the wake of the Cossack, Russia has safely emerged from her former seclusion in the northern forests. Perhaps the change of greatest intrinsic importance which took place in Europe in the last century was the southward migration of the Russian peasants, so that, whereas agricultural settlements formerly ended at the forest boundary, the centre of the population of all European Russia now lies to the south of that boundary, in the midst of the wheat-fields which have replaced the more western steppes. Odessa has here risen to importance with the rapidity of an American city.

A generation ago steam and the Suez Canal appeared to have increased the mobility of sea-power relatively to land-power. Railways acted chiefly as feeders to ocean-going commerce. But transcontinental railways are now transmuting the conditions of land-power, and nowhere can they have such effect as in the closed heartland of Euro-Asia, in vast areas of which neither timber nor accessible stone was available for road-making. Railways work the greater wonders in the steppe, because they directly replace horse and camel mobility, the road stage of development having here been omitted.

In the matter of commerce it must not be forgotten that ocean-going traffic, however relatively cheap, usually involves the four-fold handling of goods – at the factory of origin, at the export wharf, at the import wharf and at the inland warehouse for retail distribution; whereas the continental railway truck may run direct from the exporting factory into the importing warehouse. Thus marginal ocean-fed commerce tends, other things being equal, to form a zone of penetration round the continents, whose inner limit is roughly marked by the line along which the cost of four handlings, the oceanic freight and the railway freight from the neighbouring coast, is equivalent to the cost of two handlings and the continental railway freight. English and German coals are said to compete on such terms midway through Lombardy.

The Russian railways have a clear run of 6000 miles from Wirballen in the west to Vladivostok in the east. The Russian army in Manchuria is as significant evidence of mobile land-power as the British army in South Africa was of sea-power. True, that the Trans-Siberian railway is still a single and precarious line of communication, but the century will not be old before all Asia is covered with railways. The spaces within the Russian Empire and Mongolia are so vast, and their potentialities in population, wheat, cotton, fuel and metals so incalculably great, that it is inevitable that a vast economic world, more or less apart, will there develop inaccessible to oceanic commerce.

As we consider this rapid view of the broader currents of history, does not a certain persistence of geographical relationship become evident? Is not the pivot region of the world's politics that vast area of Euro-Asia which is inaccessible to ships, but in antiquity lay open to the horse-riding nomads, and is today about to be covered with a network of railways? There have been and are here the conditions of a mobility of military and economic power of a far-reaching and yet limited character. Russia replaces the Mongol Empire. Her pressure on Finland, on Scandinavia, on Poland, on Turkey, on Persia, on India and on China, replaces the centrifugal raids of the steppemen. In the world at large she occupies the central strategical position held by Germany in Europe. She can strike on all sides and be struck from all sides, save the north. The full development of her modern railway mobility is merely a matter of time. Nor is it likely that any possible social revolution will alter her essential relations to the great geographical limits of her existence. Wisely recognising the fundamental limits of her power, her rulers have parted with Alaska; for it is as much a law of policy for Russia to own nothing overseas as for Britain to be supreme on the ocean.

Outside the pivot area, in a great inner crescent, are Germany, Austria, Turkey, India and China, and in an outer crescent, Britain, South Africa, Australia, the United States, Canada and Japan. In the present condition of the balance of power, the pivot state, Russia, is not equivalent to the peripheral states, and there is room for an equipoise in France. The United States has recently become an eastern power, affecting the European balance not directly, but through Russia, and she will construct the Panama Canal to make her Mississippi and Atlantic resources available in the Pacific. From this point of view the real divide between east and west is to be found in the Atlantic Ocean.

The oversetting of the balance of power in favour of the pivot state, resulting in its expansion over the marginal lands of Euro-Asia, would permit of the use of vast continental resources for fleet-building, and the empire of the world would then be in sight. This might happen if Germany were to ally herself with Russia. The threat of such an event should, therefore, throw France into alliance with the oversea powers, and France, Italy, Egypt, India and Korea would become so many bridgeheads where the outside navies would support armies to compel the pivot allies to deploy land forces and prevent them from concentrating their whole strength on fleets. On a smaller scale that was what Wellington accomplished from his sea-base at Torres Vedras in the Peninsular War. May not this in the end prove to be the strategical function of India in the British Imperial system? Is not this the idea underlying Mr Amery's conception that the British military front stretches from the Cape through India to Japan?

The development of the vast potentialities of South America might have a decisive influence upon the system. They might strengthen the United States, or, on the other hand, if Germany were to challenge the Monroe Doctrine successfully, they might detach Berlin from what I may perhaps describe as a pivot policy. The particular combinations of power brought into balance are not material; my contention is that from a geographical point of view they are likely to rotate round the pivot state, which is

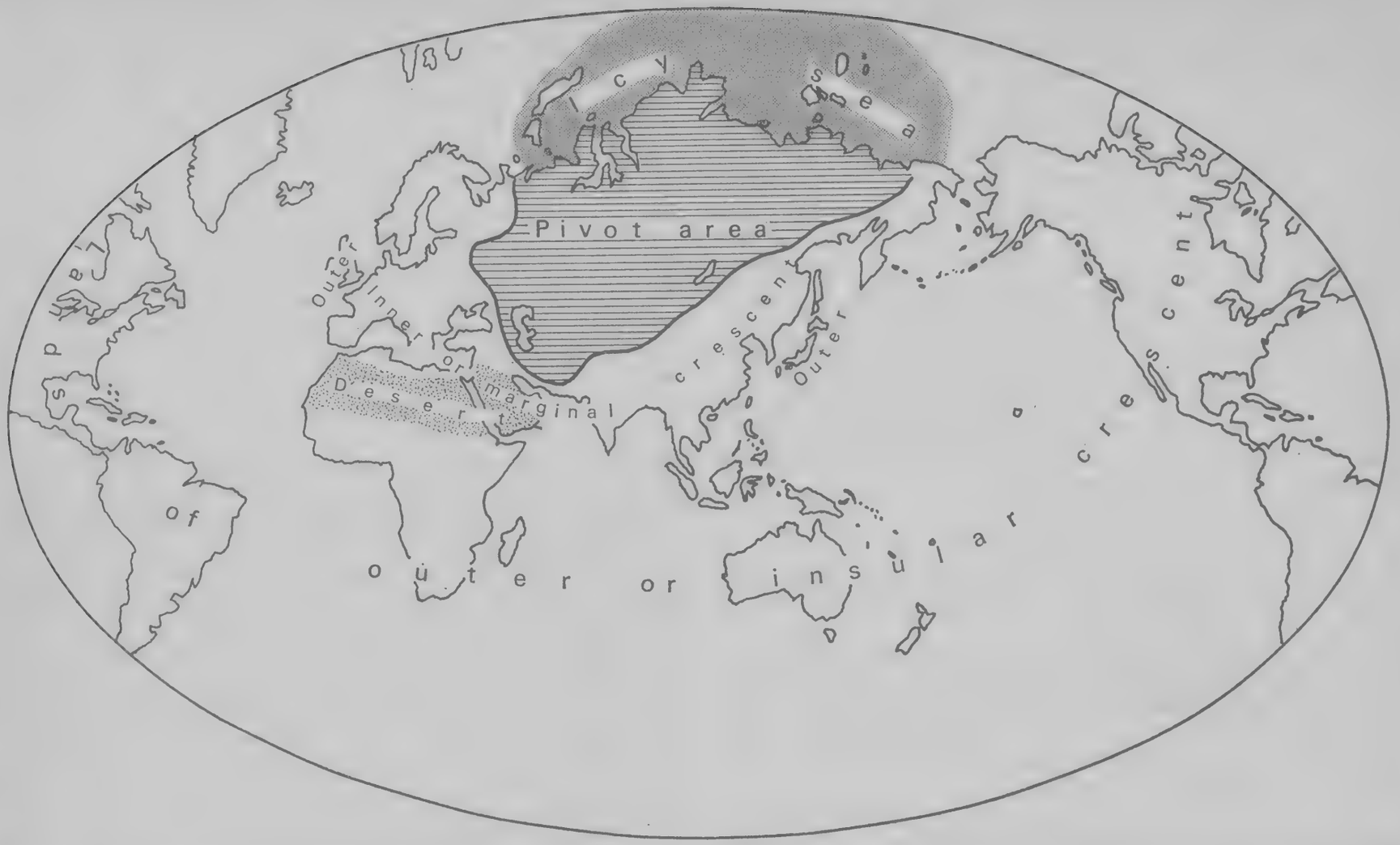


Fig. 23.5 The natural seats of power. Pivot area: wholly continental. Outer crescent: wholly oceanic. Inner crescent: partly continental, partly oceanic.

always likely to be great, but with limited mobility as compared with the surrounding marginal and insular powers.

I have spoken as a geographer. The actual balance of political power at any given time is, of course, the product, on the one hand, of geographical conditions, both economic and strategic, and, on the other hand, of the relative number, virility, equipment and organisation of the competing peoples. In proportion as these quantities are accurately estimated are we likely to adjust differences without the crude resort to arms. And the geographical quantities in the calculation are more measurable and more nearly constant than the human. Hence we should expect to find our formula apply equally to past history and to present politics. The social movements of all times have played around essentially the same physical features, for I doubt whether the progressive

desiccation of Asia and Africa, even if proved, has in historical times vitally altered the human environment. The westward march of empire appears to me to have been a short rotation of marginal power round the south-western and western edge of the pivotal area. The Nearer, Middle and Far Eastern questions relate to the unstable equilibrium of inner and outer powers in those parts of the marginal crescent where local power is, at present, more or less negligible.

In conclusion, it may be well expressly to point out that the substitution of some new control of the inland area for that of Russia would not tend to reduce the geographical significance of the pivot position. Were the Chinese, for instance, organised by the Japanese, to overthrow the Russian Empire and conquer its territory, they might constitute the yellow peril to the world's freedom just because they would add an oceanic frontage to the resources of the great continent, an advantage as yet denied to the Russian tenant of the pivot region.

Discussion following oral presentation at the Royal Geographical Society, London 25 January 1904.

Before the reading of the paper, the *President* said: We are always very glad when we can induce our friend Mr Mackinder to address us on any subject, because all he says to us is sure to be interesting and original and valuable. There is no necessity for me to introduce so old a friend of the Society to the meeting, and I will therefore at once ask him to read his paper.

After the reading of the paper, the *President* said: We hope that Mr Spenser Wilkinson will offer some criticism on Mr Mackinder's paper. Of course, it will not be possible to avoid geographical politics to a certain extent.

Mr Spenser Wilkinson: It would occur to me that the most natural thing and the most sincere thing to say at the beginning is to endeavour to express the great gratitude which, I am sure, everyone here feels for one of the most stimulating papers that has been read for a long time. As I was listening to the paper, I looked with regret on some of the space that is unoccupied here, and I much regret that a portion of it was not occupied by the members of the Cabinet, for I gathered that in Mr Mackinder's paper we have two main doctrines laid down: the first, which is not altogether new – I think it was foreseen some years back in the last century – that since the modern improvements of steam navigation the whole of the world has become one, and has become one political system. I forget the exact expression that Mr Mackinder used; I think he said that the difference was something like that of a shell falling into an enclosed structure and falling into space. I should wish to express the same thing by saying that, whereas only half a century ago statesmen played on a few squares of a chess-board of which the remainder was vacant, in the present day the world is an enclosed chess-board, and every movement of the statesman must take account of all the squares in it. I myself can only wish that we had Ministers who would give more time to studying their policy from the point of view that you cannot move any one piece without considering all the squares on the board. We are very much too apt to look at our policy as though it were cut up into watertight compartments, each of which had no connection with the rest of the world, whereas it seems to me the great fact of today is that any movement which is made in one part of the world affects the whole of the international relations of the world – a fact which, it seems to me, is lamentably neglected both in British policy and in most of the popular discussions of it, and I am exceedingly grateful to Mr Mackinder for having laid so much stress on that in his paper. Then the other point – the main point, I take it – which he has brought out is really as to the enormous importance to the world of the modern expansion of Russia. I cannot say that I am thoroughly convinced of some of Mr Mackinder's historical analogies or precedents, unless, indeed, we are to take his paper as carrying us a very long way ahead. Mr Mackinder takes us back over four hundred years, and talks of the Columbian epoch. Well, I cannot pretend to be able to go four hundred years forward; if one can go a generation forward, it is quite as much as some of us can manage. Now, these great movements of Central Asian tribes on to Europe and on to the different marginal countries may, I think, be overestimated in their importance. They have left occasional survivals of the past, but they have not left the world much richer in ideas, and very seldom represented any permanent alterations in the conditions of mankind; and they have been possible because the expanding forces of Central Asia hit upon a very much divided margin. For instance,

the movement of the Ottoman Turks, and before that the Turkish movements upon the Byzantine Empire and upon the region that had been the Byzantine Empire, invariably struck upon regions in which government was in decay or obsolescent, and most of the movements which struck upon Central Europe, the movements north of the Black Sea, struck upon Europe at a time when government was very little organised, and when the states had very little of solidarity between them. Therefore, I hold they do not afford very much parallel for the future; and I should be disposed to dwell on the counterbalancing phenomenon, which is that you have had in the west of Europe a small island, which, having attained to its own political unity, and having in the conflict for its own independence developed its sea-power, has been able to affect the marginal regions and to acquire the enormous influence which was revealed to us, a little exaggerated, perhaps, on the map which Mr Mackinder showed – the British Empire – exaggerated because it was a map on Mercator's projection, which exaggerated the British Empire, with the exception of India. My own belief is that an island state like our own can, if it maintains its naval power, hold the balance between the divided forces which work on the continental area, and I believe that has been the historical function of Great Britain since Great Britain was a United Kingdom. Now we find a smaller island state rising on the opposite side of the Euro-Asian continent, and I see no reason at all to suppose that that state should not be able to exercise on the eastern fringe of the Asiatic continent a power as decisive and as influential as that which the British Isles, with a smaller population have exercised upon Europe.

Sir Thomas Holdich: When one hears a lecture such as Mr Mackinder has just given us, so full of thought and so thoroughly well worked out, with such an amount of food for reflection contained in it, it takes a great deal of moral digestion to assimilate it, and more assurance than I possess either to criticise it, or even to discuss it. But there is just one question I should like to ask Mr Mackinder, and in co-relating the facts of geographical conditions with the history of the human race, it seems to me a not unimportant one. Mr Mackinder has told us that in the beginning of things the Mongol races all started from a centre in high Asia, spreading outwards, westwards, southwards and eastwards, finding, however, Tibet an impossible barrier in their way, and never exactly occupying India. But we must remember that before the Mongolians spread, there were other Central Asian tribes who spread equally from districts which were not so very far removed from the position which the Mongolians themselves first occupied – the Scyths and the Aryans – and that they did find their way into India. That, however, is a matter of detail. What I should like to know from Mr Mackinder is, what he considers to be the original reason of that extraordinary overflow from the country which we are disposed to consider to be the cradle of the human race, to all the different parts of the world. Was it simply the nomadic instincts of the people, a sort of hereditary compulsion which obliged them to flow outwards; or was it an actual alteration in the physical characteristics of the country in which they dwelt? We know that the physical conditions of the world alter very much from time to time, and it seems to me impossible to reconcile the idea of a great inland country, which must once have been full of a teeming population, and have supported that population, as you may say, with an abundant power of agricultural wealth – that under such conditions a people should have had a desire to spread out and to wander forth into other parts of the world, seeking for they knew not what. I fancy, myself, that one of the great reasons, one of the great compelling reasons, for all these migrations really has been a distinct alteration in the physical conditions of the country. That is a point which seems to me to be rather important when we are discussing a subject like the present one, which brings the conditions of geography to bear on the facts of history. There is just one other little matter which was referred to somewhat doubtfully by Mr Mackinder to which I might refer. He pointed to South America as a possible factor in that outer belt of power which was to bring coercion to bear on the inner power pivoting about the south of Russia. Now, from what I have seen lately, I have not the least doubt that that will be the case. The potentiality of South America as a naval power I look upon as very great. I believe that in the course, say, of the next half-century, in spite of the fact that just now Argentina has sold two ships to Japan, and Chile has sold a couple of ships to us – in spite of that fact, there will be an increase of naval strength in South America, resulting from purely natural causes, for the defence of her own coast and the protection of her own traffic, which will be only comparable to the extraordinary development

which we have seen during the last half-century in Japan. This seems to me certainly to be one of the factors, if we are to look forward, with which, in the future naval politics of the world, we shall have to reckon.

Mr Amery: I think it is always enormously interesting if we can occasionally get away from the details of everyday politics and try to see things as a whole, and this is what Mr Mackinder's most stimulating lecture has done for us tonight. He has given us the whole history and the whole of ordinary politics under one big comprehensive idea. I remember when I did Herodotus at the university, he made the whole of history base itself upon the great struggle between the east and the west. Mr Mackinder makes the whole of history and politics base themselves on the great economical struggle between the great inside core of the Euro-Asiatic continent and the smaller marginal regions and islands outside. I am not sure myself that these two struggles are not one and the same, because now we have discovered that the world is a sphere, east and west have only become relative terms. I would criticise one thing Mr Mackinder said when he described Russia as the heir of Greece. It was not the ancient heir of Hellenic Greece, but of Byzantium, and Byzantium was the heir of the old Oriental monarchies with the Greek language and a tinge of Roman civilisation thrown over it. I should like to go back, if I might, for a moment to this geographical economic foundation on which Mr Mackinder built the framework of his lecture. I think I would conceive the thing somewhat differently. There are, to my mind, not two, but three economico-military forces. If we begin with the ancient world, you have the broad geographical division into the 'steppes' of the interior, the rich marginal land suitable for agriculture and the coast, and you have, corresponding with these, three economical and three military systems. There is the economical and military system of the agricultural country, the system of the coast and seafaring people, and the system of the steppes; each had its peculiar weaknesses and its peculiar sources of strength. The strongest in many ways was the marginal and agricultural state. There you got the great solid military empires, your Egyptian, your Babylonian, your Roman Empire, your large armies and citizen infantry, your great development of wealth. But these contained certain elements of weakness. Their own prosperity or the defects of their form of government would lead ultimately to slothfulness and weakness. Now, outside those you had two other systems. You had the steppe system, whose military strength lay, firstly, in its mobility, and, secondly, in its inaccessibility from the slower-moving agricultural powers. As regards the supposed 'hordes' of invaders which came from the interior, I do not myself believe there ever were those very large hordes and large populations in the interior. The fact is this, the steppe populations were small then as now, but from the fact of their mobility the heavier and slower military armies could not successfully attack them. In ordinary times, when the agricultural states were strong, the people of the steppes simply ran away from them, and the others found it too much trouble to conquer them. You remember the difficulty the Roman legions had with the Parthians; and I think we can find a very much more recent example of the difficulty a civilised state finds in conquering a steppe power. Only a short time ago the whole of the British army was occupied in trying to coerce some 40,000 or 50,000 farmers who lived on a dry steppe-land. That photograph Mr Mackinder was showing reminded me exactly of what you could have seen not so many months ago in South Africa – I mean, that picture of wagons crossing the river was, except for the shape of the roof over the waggon, exactly like a picture of a Boer commando crossing a drift. We had the same difficulty in coercing them that all civilised powers have had with steppe people. Now, whenever the civilised powers on the marginal countries have grown weak and have allowed small hired armies to do their work, they have got into difficulties, and that is where, it seems to me, the strength of the steppes has always come in. There is no great economic strength at bottom, but the fact that they could retire into their inaccessible wilderness, and come upon the others in times of their weakness, gave the steppe peoples their power. Then there is the third system, that of the maritime coast peoples: they had even less pure military strength, but they had the greatest mobility – the mobility, I mean, of the Vikings or the Saracens when they ruled the Mediterranean, and the Elizabethan Englishmen when they harried the Spanish Main. Coming to more modern times, there has been a certain further change in the agricultural conditions, and the development, out of the old agricultural states, of the modern industrial state. Then I would also notice that many countries which were steppe became agricultural and industrial. You have that, and

you have also the fact that very rarely in history do you get any state rising to great power by one system alone. The Turks began by being the people of the steppes, and came down and swept over Asia Minor; they then formed a regular military power, and conquered the great Turkish Empire; lastly, for a period they became the leading naval power in the Mediterranean. In the same way, you find the Romans, in order to beat the Carthaginians, became a sea-power as well as a land-power; and, in fact, for a power to be great it must have both these elements of strength. The Romans were a great military power with the marginal region as their base and with sea-power behind them. We ourselves have always had as a base the industrial wealth of England. The Russian Empire, which covers the great steppe region, but is no longer in the hands of the old steppe people, is really a portion of the agricultural world, economically, which has conquered the steppe and is turning it into a great agricultural industrial power, and therefore giving a power which the pure steppe people never possessed. Mr Mackinder referred to the fact that it is only within the last century that the agricultural races have occupied and populated the southern steppe of Russia proper. They are doing the same thing in Central Asia; in fact, the old steppe people are being squeezed out altogether, and you get, coming closer and closer together, two leading industrial-military powers, the one radiating out from a continental centre, and the other beginning from the sea, but gradually going farther into the continent in order to have the big industrial base which it requires, because sea-power alone, if it is not based on great industry, and has a great population behind it, is too weak for offence to really maintain itself in the world struggle. I do not intend to make many more remarks, but there is just one point – a word of Mr Mackinder's suggested it to me. Horse and camel mobility has largely passed away; and it is now a question of railway mobility as against sea mobility. I should like to say that sea mobility has gained enormously in military strength to what it was in ancient times, especially in the number of men that can be carried. In the old days the ships were mobile enough, but they carried few men, and the raids of the sea-people were comparatively feeble. I am not suggesting anything political at the present time; I am merely stating a fact when I say that the sea is far better for conveying troops than anything, except fifteen or twenty parallel lines of railway. What I was coming to is this: that both the sea and the railway are going in the future – it may be near, or it may be somewhat remote – to be supplemented by the air as a means of locomotion, and when we come to that (as we are talking in broad Columbian epochs, I think I may be allowed to look forward a bit) – when we come to that, a great deal of this geographical distribution must lose its importance, and the successful powers will be those who have the greatest industrial basis. It will not matter whether they are in the centre of a continent or on an island; those people who have the industrial power and the power of invention and of science will be able to defeat all others. I will leave that as a parting suggestion.

Mr Hogarth: As the hour is rather late and the temperature rather low, I will not take up your time with any very lengthy remarks. We certainly have had a wonderfully suggestive paper, and I think it is neither necessary to advise the reader of the paper nor anyone who had listened to it to try and think imperially. I would only ask Mr Mackinder, when he replies, to make me certain about one point. Does he really mean to imply – I think it is an interesting fact if he meant to establish it – that the state of things which is coming to pass in this inner pivot land will be entirely different to anything that has been seen there before? That is to say, something like a stationary state of things has been brought about, and the country is being developed, till it will even be able to export its own products to the rest of the world; and therefore we are never to see again the state of things that has existed all through ancient history in that a great central region which has continually sent its populations down into the marginal countries, while the marginal countries have sent back to it their influences of civilisation, each operating in turn upon the other. The only other observation I would like to make is to reinforce Mr Amery's objection to Mr Mackinder's Graeco-Slav. I am afraid I cannot accept that division of civilisation between the Greek and the Roman. So far as Russia can be called a civilised country at this moment, it has, I think, not been civilised by the Orthodox Church; in fact, I have yet to learn of any civilising influence exerted by the Orthodox Church on a great scale. Its civilisation is far more due to the social culture which was introduced by Peter the Great, and that was more Roman than Greek. But it is to my first question I should like Mr Mackinder to give a clear answer. I should like to know what he seriously anticipates is going

to be the effect on the world of this new distinction between the marginal and the central pivot lands.

Mr Mackinder: I have to thank all the speakers for dotting my *i*'s and crossing my *t*'s. I am delighted to find my formula work so well. I do mean exactly what Mr Hogarth says; I mean that for the first time within recorded history – and this is in reply to Sir Thomas Holdich as well – you have a great stationary population being developed in the steppe lands. This is a revolution in the world that we have to face and reckon with. I doubt very much, and there I agree with Mr Amery, whether the numbers who came from the heart of Asia were very great. It seems to me quite as he puts it, and that their mobility was of the very essence of the whole thing. A small number of people coming from the steppe lands could do many things, given relative mobility as compared with the agricultural population. With regard to Sir Thomas Holdich's inquiry as to what should send them forth, Sir Clements Markham has pointed out that the nomads did not pour forth once only. I dealt with the fact that for a thousand years the nomadic peoples came through Russia. I fail to see that, when you have this constant succession of descents upon the marginal lands, you are called upon to ask for any special physical change to explain it. All the accounts we have from the time of the earliest Greeks describe the drinkers of mares' milk, and picture for us the nomadic mode of life; therefore I start with the fact that these peoples were nomadic and remained nomadic through two thousand years, and I do not see any evidence that we need either to call in any great physical change or yet to assume any great settled population. As far as I can see, Sven Hedin refuses the idea that you must necessarily ask for a great change of climate in order to explain the existence of the remains in Central Asia. You have powerful winds and much sand, and from time to time the sand is swept over hundreds of miles across the desert. The sand determines the flow of the rivers and the position of the lakes, and some great storm diverting a river into another course would no doubt suffice to ruin a town abandoned by the water. The mere fact that there were nomads, and that there were rich countries to be plundered, seems to me to be almost sufficient for my theory. In the future, I think, you are bound to have different economic provinces, one based mainly on the sea, and the other on the heart of the continent and on railways. I do not think Mr Amery has allowed sufficiently for the fact that the very largest armies cannot be moved by means of a navy. The Germans marched nearly a million men into France; they marched, and used the railways for supplies. Russia, by her tariff system and in other ways, is steadily hastening the accomplishment of what I may call the non-oceanic economic system. Her whole policy, by her tariff system, by her break of gauge on her railways, is to separate herself from external oceanic competition.[3] With regard to the basis of sea-power in industrial wealth, I absolutely agree. What I suggest is that great industrial wealth in Siberia and European Russia and a conquest of some of the marginal regions would give the basis for a fleet necessary to found the world empire. Mr Amery's way of putting the three groups of powers is slightly different from mine, but it is essentially the same. I ask for an inner land mobility, for a margin densely populated and for external sea forces. It is true the camel-men and the horsemen are going; but my suggestion is that railways will take their place, and then you will be able to fling power from side to side of this area. My aim is not to predict a great future for this or that country, but to make a geographical formula into which you could fit any political balance.

There was a point with regard to the Graeco-Slav: in the sense in which Mr Hogarth and Mr Amery have taken me, I agree with them, but after all I cannot help feeling that Christianity fell on two very different soils – the Greek philosophic and the Roman legal – and that it has in consequence differently influenced the Slav and the Teuton. However, that is a mere incident, and if I qualify my statement by speaking of the Byzantine, I shall then get near to what Mr Amery asks, and I think I shall do away with the necessity of introducing the example of Rome which Mr Hogarth brought forward. As regards the potentialities of the land and of the people, I would point out that in Europe there are now more than 40 million people in the steppe land of Russia, and it is by no means yet densely occupied, and that the Russian population is probably increasing faster than any other great civilised or half-civilised population in the world. With a decreasing French population, and a British not increasing as fast as it was, and the native-born populations of the United States and Australia coming nearly to a standstill, you have to face the fact that in a hundred years 40 million people have occupied but a

mere corner of the steppe. I think you are on the way to a population which will be numbered by the hundred million; and this is a tendency which you must take into account in assigning values to the variable quantities in the equation of power for which I was seeking a geographical formula. The point with regard to Korea and the Persian Gulf which was put by Mr Spenser Wilkinson exactly illustrates my correlation of the Far Eastern, Middle Eastern and Near Eastern questions. I represent these as being the present temporary form of the collision between the external and internal forces acting through the intermediate zone, which is itself the seat of independent forces. I quite agree that the function of Britain and of Japan is to act upon the marginal region, maintaining the balance of power there as against the expansive internal forces. I believe that the future of the world depends on the maintenance of this balance of power. It appears to me that our formula makes it clear that we must see to it that we are not driven out of the marginal region. We must maintain our position there, and then, whatever happens, we are fairly secure. The increase of population in the inner regions and the stoppage of increase in the outer regions may be rather serious; but perhaps South America will come in to help us.

The President: I confess I have been entranced by Mr Mackinder's paper, and I could see by the close attention with which it was listened to by the audience that you all shared my feeling in that respect. Mr Mackinder has dealt with the old, old story from the very dawn of history, the struggle between Ormerzd and Ahriman, and he has shown us how that struggle has continued on from the very dawn of history to the present day. He has explained all this to us with a brilliancy of description and of illustration, with a close grasp of the subject and with a clearness of argument which we have seldom had equalled in this room. I am sure you will all with me give a unanimous vote of thanks to Mr Mackinder for his most interesting paper this evening.

Notes

1. Read at the Royal Geographical Society, 25 January 1904.

2. This statement was criticised in the discussion which followed the reading of the paper. On reconsidering the paragraph, I still think it substantially correct. Even the Byzantine Greek would have been other than he was had Rome completed the subjugation of the ancient Greek. No doubt the ideals spoken of were Byzantine rather than Hellenic, but they were not Roman, which is the point.

3. The Russian customs ring is, of course, so placed as to add to the pivotal area for economic purposes considerable sections of the marginal lands, although not of the oceanic coasts. – J. H. M.

THE HEARTLAND DEBATE

H. de Blij

MACKINDER'S hypothesis immediately became one of the most intensively debated geographical ideas of all time. He himself joined in the discussion, modifying the concept in some minor areas, but arguing, as late as 1943, that it 'is more valid and useful today than it was either twenty or forty years ago'. Upon and beneath the Heartland, Mackinder asserted, lies a store of rich soil for cultivation and of ores and fuels for extraction, 'the equal – or thereabouts – of all that lies upon and beneath the United States and the Canadian Dominion'.

Among those geographers who disagreed was N. J. Spykman, whose criticism was directed at two weak points in the Heartland concept, both of which are reflected by the quotations in the previous paragraph. In the first place, Spykman felt, Mackinder had overemphasised the potentialities of the Heartland. In many respects, the real power potential of Eurasia lay in what Mackinder had called the 'Inner or Marginal Crescent'. This region Spykman called the 'Rimland', and he countered with his own theory:

> Who controls the Rimland rules Eurasia;
> Who rules Eurasia controls the destinies of the world.

Although this answer came some decades after Mackinder had stated his final hypothesis and this had the benefit of a longer period of history, Spykman's basic objections were as true in 1904 and 1919 as they were in 1944. Furthermore, Spykman pointed out relevantly that history involving the Heartland never really was a simple landpower–seapower opposition; some of the states in the Rimland were at one time or another allied with Russia; at other times Russia and Great Britain had been allied in opposition to a Rimland power.

Mahan, Mackinder and Spykman all discussed the Heartland concept in terms of land and sea power. But air power has gained

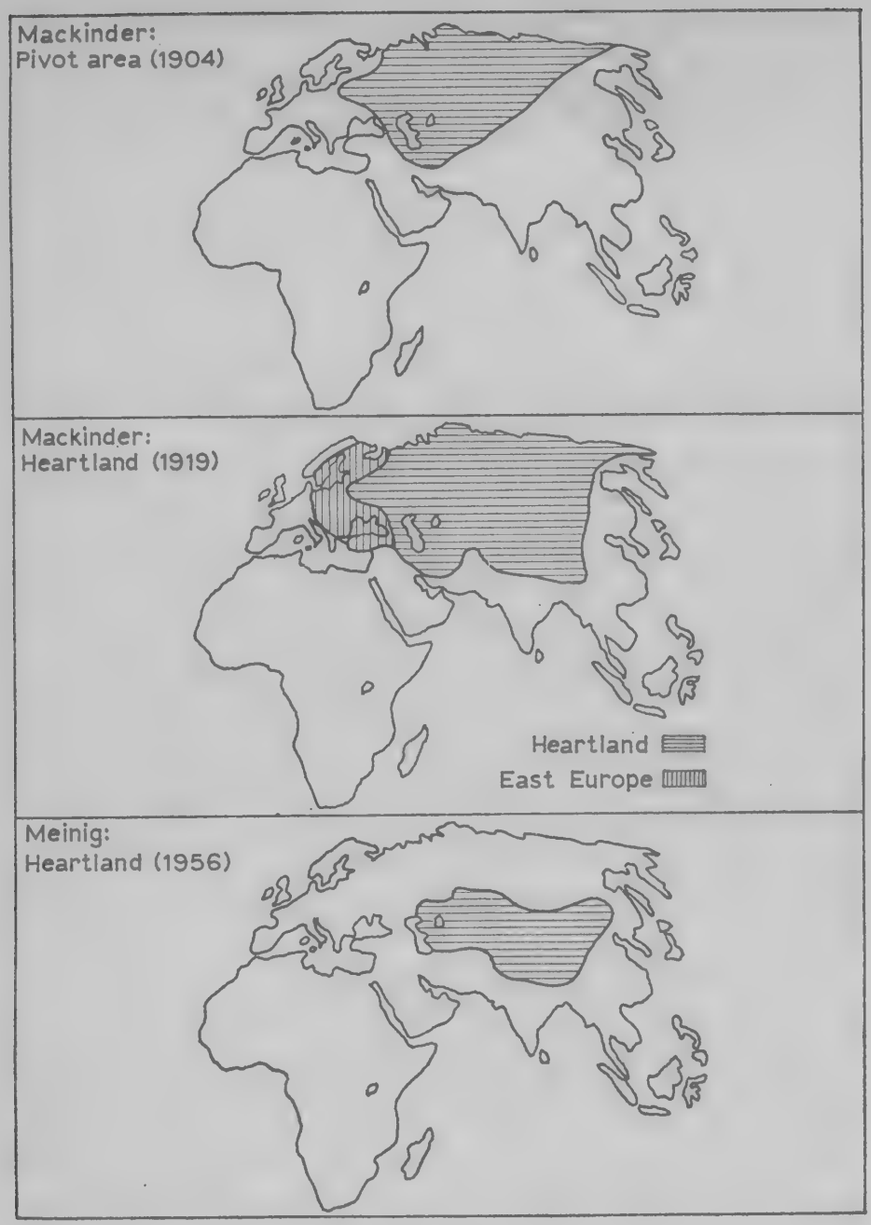


Fig. 23.6

Sources:
H. J. Mackinder, 'The Geographical Pivot of History', *Geogr. J.*, vol. 23 (1904) 421–44.
H. J. Mackinder, *Demographic Ideals and Reality* (London, 1919).
D. W. Meinig, 'Heartland and Rimland in Eurasian History', *Western Pol. Q.*, vol. 9 (1956) 553–69.

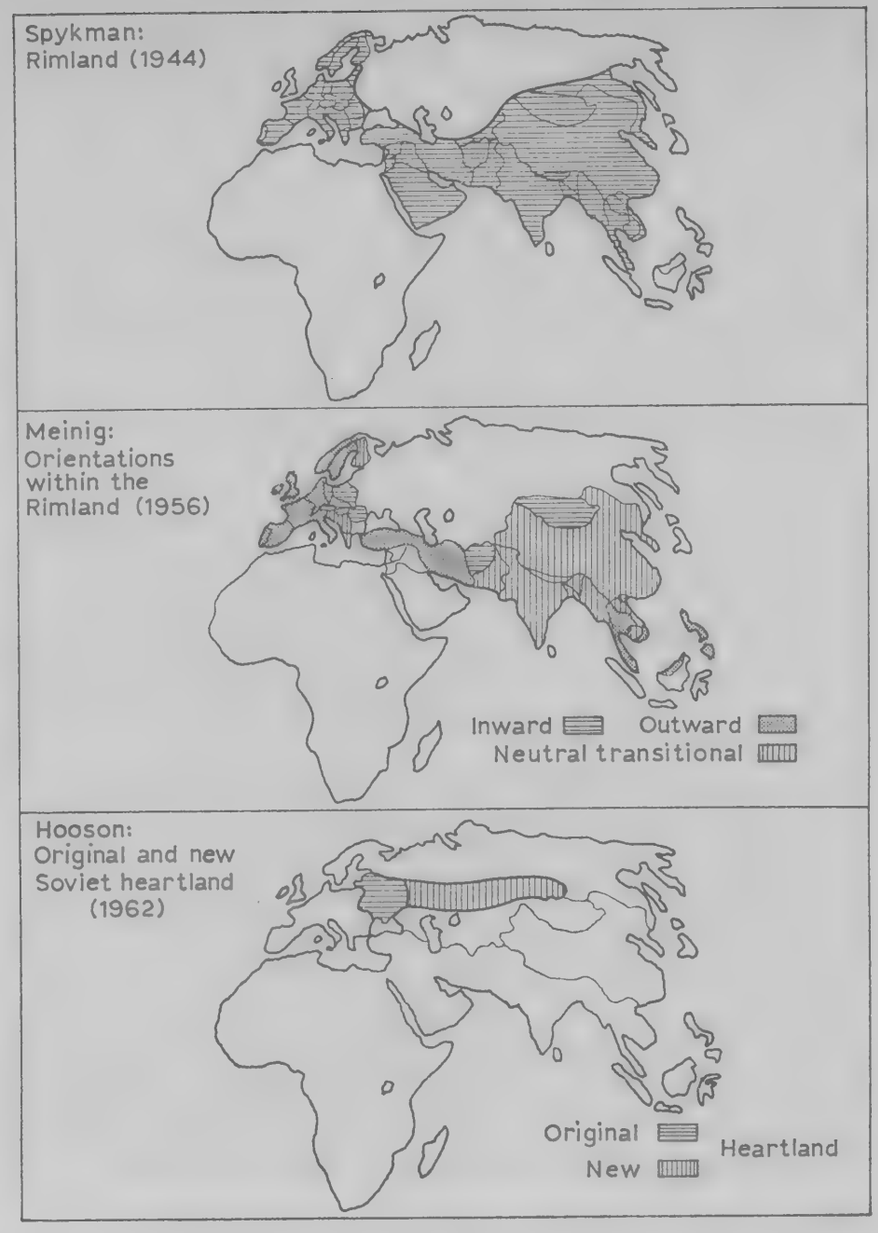


Fig. 23.7

Sources:
N. J. Spykman, *The Geography of the Peace* (New York, 1944).
D. W. Meinig, op. cit.
D. J. M. Hooson, *A New Soviet Heartland?* (Princeton, 1964).

to such an extent in the world strategic picture that their theories may be completely invalid today. Mackinder in 1943 was not convinced that this could happen:

> Some persons today seem to dream of global air power which will 'liquidate' both fleets and armies. I am impressed, however, by the broad implications of a recent utterance of a practical airman: 'Air power depends absolutely on the efficiency of its ground organisation.' . . . It can only be said that no adequate proof has yet been presented that air fighting will not follow the long history of all kinds of warfare by presenting alternations of offensive and defensive tactical superiority, meanwhile effecting few permanent changes in strategical conditions.

One strategist who was indeed convinced was A. P. de Seversky, who in 1950 published the view that the land and sea forces of the world have been subordinated to the greater power of the air. De Seversky advocated the development of massive air superiority as the prime goal for American policy-makers. Small wars (such as that in Vietnam) are a useless sapping of United States energies (while costing the adversaries almost nothing); overseas bases are costly luxuries that should be abandoned. United States superiority should be so complete that it would be possible to contain any enemy by stating exactly the line across which he may not step; any such aggression will result in the total annihilation of the enemy by the massive retaliatory forces of superiority that air power has brought the United States. De Seversky looks at the world map in a different way – not conventionally as did Mackinder and Spykman, but on the azimuthal equidistant projection centred on the North Pole, showing quickly the areas of air dominance of the United States and the U.S.S.R.

De Seversky's map leads us to another major criticism of Mackinder's Heartland concept, and to a most comprehensive statement of the assets and shortcomings of the idea. In 1955, A. R. Hall pointed out that Mackinder had overlooked, in his world scheme, the existence of another heartland in Anglo-America which was and remains capable of occupying as important a place in planetary affairs as the one in Eurasia.[1] Peculiarly Mackinder mentions this heartland in his 1943 paper quoted above, but he fails to follow up the apparently obvious consequences.

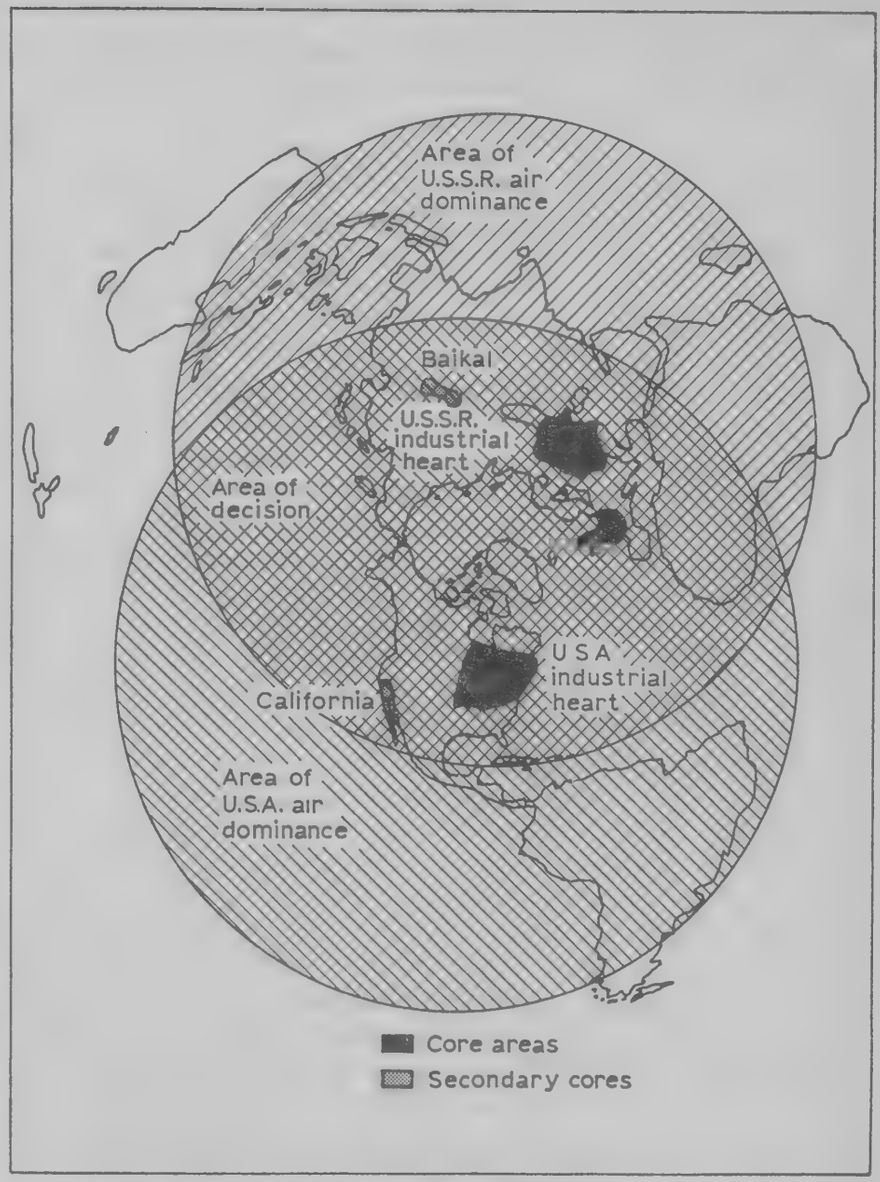


Fig. 23.8 The airman's view: de Seversky

Source: A. de Seversky, *Air Power: the Key to Survival* (New York, 1950).

De Seversky emphasises this weakness, and Hall views it as one of the major faults in the thesis.

Another weakness in the Heartland concept relates to its date of publication, 1904. Although time was to substantiate Mackinder's assumptions regarding the Heartland to a degree even he himself could probably not have imagined, it must not be forgotten that the original statement of the thesis was based upon data available prior to 1904. If, indeed, Mackinder's hypothesis constituted a 'geographical formula into which you could fit any political balance', as Mackinder himself asserted, and if the pivot area gave so much power to the possessor, why was Russia in 1904 not more of a world power than it was?

Perhaps the most frequently levelled charge against the Heartland concept is that it constitutes an oversimplification of history and a dangerous determinist prediction. Mackinder's greatest fault, revealed especially in his 1943 statement, seems to have been his inability to recognise the role of changing technology in the present-day political world. In 1956, D. W. Meinig published an article that sought to tighten the definitions involved in Mackinder's original thesis and in subsequent discussions of it.

Meinig stressed that Mackinder as well as his most important critic (Spykman) had focused their attention upon the particular geopolitical context of their time. This, he argued, resulted in an inflexibility in their concepts, so that their relevance would become increasingly historical and less and less applicable to the dynamic changes of the present-day world. If such terms as Heartland and Rimland are to have maximum utility, 'applicable beyond any momentary context of strategic patterns, those definitions must become specific in concept yet flexible in historical-spatial use. . . . Our definitions of heartland and rimland must . . . be rooted in cultural, or, I should like to term them, functional criteria.' Thus Meinig retraces the criteria on the basis of which Mackinder established his Heartland, pointing out that hydrographic considerations were paramount: he then defines a new heartland (see Fig. 23.6), calling it more stable, yet functional. The qualities of this heartland include broadly similar physical conditions and basically similar cultures, an indisputable interior location (i.e. nodal position), and the nexus of all the historic land routes interconnecting the several rimland areas of China, India, the Levant and Europe.

More significant and useful, perhaps, is Meinig's re-evaluation of the Rimland. He points out that both Mackinder and Spykman, though they gave different names to this region, viewed this area as the natural realm mainly of the sea powers of the world. But what of the states forming part of the Rimland? Their own orientation is of paramount importance and changes with time. Some Rimland states have been closely tied to the Heartland power in terms of ideology and trade during one period of recent or distant world history; at other times the orientation may have been outward and maritime. Within the past two decades Yugoslavia's orientation as a Rimland state has changed very much from 'inner' (towards the Heartland power) to 'outer'. On a functional basis, thus, Meinig suggests that a *continental* rimland and a *maritime* rimland should be recognised. In this scheme, certain states might perhaps be considered to be transitional or neutral. There are signs that Pakistan, for example, is moving from an 'outer' orientation towards an 'inner' one: like Yugoslavia, it may go through a neutral phase.

A very recent contribution to our thinking regarding the Heartland is that of D. J. M. Hooson. In attempting to define the core of the Soviet Union, and to identify the real heart of that state, Hooson examines a number of regions which form part of the whole Heartland identified by Mackinder. He uses a set of most relevant criteria: the Soviet Union, as today's Heartland power, is sustained to a greater degree by certain parts of the Heartland than by others. In Mackinder's day the focus was well to the west – where is it today? What is the scale of contribution to the national economy as a whole of certain parts of the Heartland? What is the rate of population growth, and the rate of urbanisation and urban development? Where are the bulk of the *accessible* resources? In which areas is there that critical economic specialisation involving both agriculture and industry? Is there historical association and cohesiveness? What are the ethnic conditions? What emerges from Hooson's examination is a recognition of the established 'European' core around Moscow and to the Black Sea – the centre of Russian and Soviet power for many decades – and a significant eastern zone, the Volga–Baikal region. It is this eastern zone which Hooson views as critical: 'what happens to and in this zone, between the Volga and Lake Baykal, might well have a decisive effect on the fortunes of the Soviet Union as a world power . . .'.

If, indeed, this region becomes the focus of the Soviet state – it is presently outstripping the established Moscow core in terms of rate of development – the point of gravity of the Heartland will be nearer to that suggested by Meinig. Just as Meinig has pointed out that the term Rimland constitutes an oversimplification, so Hooson indicates that the Heartland really consists of a long-established core, an expanding and vital eastern triangle, and a marginal zone whose contribution to the Heartland power's strength is yet relatively insignificant. In any view of the role of the Heartland in world politics, this must be understood.

It is obviously difficult, in thinking about these matters today, to escape strategic and geopolitical problems involved in the emerging ideological struggle between the Communist and non-Communist world. The views of several of those authorities mentioned above have as their ultimate aim the containment of any power that finds itself in a position to make an attempt to gain power over the entire world, such as, according to Mackinder, the present-day Heartland state might be. None of these views openly states a precedure for the acquisition of such world power. However, the world has already seen the implementation of the global strategic views of one group of geopoliticians: those of inter-war Germany, notably Haushofer. Although Haushofer's actual influence upon German policy after 1933 may not have been as great as some have supposed, he influenced many decision-makers in Germany through his teachings at the university. Certainly the decisions of 1933–9 were in accord with his ideas, but several war moves (especially the attack upon the Soviet Union in 1941) were totally against his views. Nevertheless, as Pounds has said, 'Haushofer contributed to that climate of opinions which regarded aggression and conquest as a natural and proper role for a powerful State.'

Whether geopolitics is actually a political view of political geography while other branches of political geography focus upon the geographical phenomena in the field remains a matter open for debate. Few political geographers today are willing to grant that geopolitics is a science in the same terms that are normally applied to political geography, but this is related to the methods used by German geopoliticians within recent time. Mahan and Huntington described themselves as geopoliticians, but that was before the fatal rise of the German school in this field. Probably some decades will pass before the term comes into general use again to mean global strategic views.

Note

1. A. R. Hall, 'Mackinder and the Course of Events', *Annals Amer. Assoc. Geogr.*, vol. 65 (1955) 109–26.

A LOCAL PERSPECTIVE ON BOUNDARIES AND THE FRONTIER ZONE: TWO EXAMPLES FROM THE EUROPEAN ECONOMIC COMMUNITY

J. W. House

CONTEXT AND PERSPECTIVE

MANY political geographers, who in recent years have felt a need to re-establish the reputation of their discipline after the damage which it has suffered through confusion with an aberrant form of geopolitics, have attempted to substitute a more cautious, empirical approach to specific territorial problems of political organisation for the formulation of vast, often untested generalisations characteristic of an earlier period. This reorientation, which preceded that currently in progress throughout other branches of human geography, has taken the form of a focusing of attention upon case studies rather than upon general principles about states and their interrelationships.

The study of frontiers and boundaries is undergoing particular revaluation at the present, requiring a much wider, more sophisticated perspective than that established by earlier generations of security-conscious soldier–statesmen. In the Western world at least, where the tide of aggressive nationalism is temporarily receding, economic and social relations between states are seen to be more complicated, and yet ultimately more fundamental than strategic issues, while absolute sovereignty is being modified in some areas by the emergence of the supranational principle. At the same time the perspectives of both political and general geographers are being directed increasingly to the study of margins of widening diversity, establishing new techniques and seeing the possibilities of integration and intellectual interchange

from the study of these perimeters. This has important implications for the study of frontiers and political boundaries, the study of whose principles has been in danger of ossification for some decades.

In physical geography the glaciologist and periglacial morphologist are actively concerned with the ice-margins, as is the biogeographer with the ecological limits of cultivation and the land–sea margin. Economic geographers are tentatively exploring the territorial implications of the concept of marginal utility: for example, the economic profitability of marginal land uses, the limits of the Central Business District or the rural–urban margin. These boundary studies converge with those of the political geographer in the definition and investigation of territorial problem areas within states, where resources may be few, populations declining and where politically augmented isolation from the economic locus of the state adds a marked deterrent to development. As August Lösch once forecast, the frontier zone often gives rise to a depressed area, the consequences of international policy having underlined an already underprivileged status.

In testing the widening context of boundary study within general and regional geography, it seems particularly appropriate to examine these relationships initially at local level, by topographical-scale investigations into the realities of life among communities within the frontier zone flanking an international boundary. The problems of border communities provide a useful subject for comparative study in differing geographical circumstances, and the limited generalisations which emerge from community-level study establish a sound base for a wider interpretation of boundaries and the significance of the frontier zone.

CLASSIFICATION AND CASE STUDY

The study of political boundaries traditionally has progressed in two scarcely related directions, by statement and testing of general principle leading to a reasoned classification, often as a major objective, and, increasingly, by the examination and comparison of individual case studies of entire borderlines. This latter more empirical approach has flourished from the analysis of disputed boundaries, and as the result of work political geographers have achieved in the aftermath of two world wars. The development of case studies has been ill co-ordinated, however, and there has been little relation to the body of formal principles earlier put forward. The methodology of the study of boundaries has lagged behind the fruitful, diverse but ill-related range of individual research investigations, most of them concerned with contested frontiers.

The lack of relation of most case studies to earlier principles and attempted classifications arises partly from the unsatisfactory nature of many generalisations hitherto attempted on the subject of political boundaries. It is also due both to the inappropriateness of at least some of the latter to the modern world, particularly to most of its peaceable areas, and to the failure to relate many of the principles formulated about boundaries to the field of general and regional geography. Defence, security and readjustment to treaty changes of boundary do not represent the whole of this subject, nor are they particularly appropriate to the present state of world boundaries.

Much orthodox work on principles of boundary study mirrored the merits and defects of contemporary work in general geography. An earlier focus, almost a fixation, on the merits of natural features for the location of political boundaries, for example, had its counterpart in the early twentieth-century hypothesis of the natural region, put forward by A. J. Herbertson and others, but now also out of fashion among most Western geographers. Later classifications of boundaries, in terms of origins, character and extent of success in operation, the latter often subjectively determined, were more sophisticated but only gradually was there a widening interpretation of the non-military and non-strategic importance of these characteristic limits of sovereignty. It was as appropriate in the latter nineteenth and early twentieth century to have a perspective on the defence of the nation-state and the security of its boundaries, both in Europe and in the areas of expanding colonialism, as it would be inappropriate to think exclusively in these terms today.

An excessive emphasis seems traditionally to have been placed on the classification of boundaries largely as an end in itself, rather than as an intermediate hypothesis for the understanding and comparison of the character and the problems of the frontier contact-zones between distinctive sovereignties. That terms must be defined, the range of boundary categories identified, described

and their outline effects assessed no one would deny, any more than that grammatical structure, with its irregular verbs, is an essential preliminary to talking meaningfully in a foreign language. Such a tool of analysis, a card index for data, cannot, however, remain an end in itself, any more than the delineation of the boundaries of other types of regions can stand as the prime purpose of study in general geography. 'Peu importe le marge, c'est le cœur qu'il faut avant tout considérer,' wrote Febvre, and Richard Hartshorne put this recently into a political context, reminding us that it is the political areas being bounded rather than their boundaries that should be our major focus of attention.

It may reasonably be concluded that the classification of boundary lines has produced valuable, if occasionally misleading, working concepts but that few of these concepts have progressed beyond a very generalised formulation and some are tied to outdated conceptions of geography. Attempts to categorise more sharply and in greater detail have often broken down on the fact that individual cases of similar boundary types occur in such differing geographical contexts that their identity is with the context, and not with the type of boundary related to natural or other features of the landscape. The possibility of universally applicable generalisations about, for example, river-line boundaries or lines of reference is more likely to emerge from the limited generalisations based upon individual case studies, than from those premature categorisations, once beloved by and indeed more appropriate to the military school of boundary-defenders and delineators.

NORMAL BOUNDARIES: THE NEED FOR AN ECONOMIC AND SOCIAL PERSPECTIVE

A survey of world boundaries today confirms that the vast majority are uncontested and mostly have prime economic and social functions, few are controversial and on even fewer is there direct military confrontation or is active revisionism being canvassed. A first group concerns states whose very existence is unacceptable to some neighbouring state which would therefore regard all boundaries of the offending state as equally abhorrent. The situation on the borders of Israel, the conflict between the partitioned states of Korea or Vietnam or the disputed frontier zone of the South Arabian Federation against the Yemen are all of this type. The ill-defined borders of the new state of Somalia and the migratory habit of the border peoples indicates a very old type of transgression but an unusual feature of conflict in the modern world. The most direct confrontations in recent times have been along the Himalayan borderland between India and the Chinese People's Republic or in the short-lived border war between Morocco and Algeria in the north-western Sahara. The Oder–Neisse line remains unacceptable to the German Federal Republic as the eastern boundary to any future German state, and there are unsolved but quiescent boundary problems in Latin America, particularly in the High Andes.

These examples of problem-dominated boundaries are nevertheless surprisingly few in number, and rarely resemble the classical nineteenth-century border confrontations between aggressive nationalisms. The defensive qualities of particular types of boundary feature are less applicable to strategic thinking in a nuclear age in any case, while the growth of supra-national concepts in both the Communist and Western world implies new and ideological bases for revaluation of boundaries and the significance of the frontier zones within which they lie. At the same time the rapid evolution of new post-colonial nation-states throughout the tropical world may presage further boundary conflicts once the forces of independent nationalism are directed to the artificial nature of those many frontiers hastily delineated in the late nineteenth-century scramble for colonies.

Minghi has commented on the few studies yet made, and the inadequate attention so far given, to normal boundary contacts between states. Such contacts logically concern economic and social linkages, the nature and extent of trade, the exchange and traffic in goods and services, and the trans-border movements of people, both journey-to-work movement, tourism or permanent settlement. These contacts must be viewed dynamically, since foreign economic policies change, and trade and migrations or price differentials and economic advantage alter at the borderline. The study of such subjects began with the theoretical economists, notably August Lösch, who considered the distortions introduced into circulation fields of goods and services by the interposition of an inhibiting feature such as a political boundary line.

The field opened up by the theoretical economists has been

little worked as yet by the political geographer, though its utility and contemporary importance are very great. At all levels in world economic society there is strong consciousness of the great disparities in resources, both natural and human, and in living standards, between one area and another. The relationships between welfare policies, foreign policy and trade are ever closer, not only between power blocs themselves but also within them and between power blocs and the individual states of the developing world.

In individual states economic imbalance is frequently focused within the frontier zone, as the regional variations in resources, prosperity and economic growth tend to accentuate contrasts in marginal territories, not a few of which are located along or near to the international boundary and far from the developing locus of economic or political power in a state.

Regional policies of governments are notably concerned with the depressed areas of the economy, the areas of slow economic growth or stagnation, and also with the need to define and implement networks of economic and administrative regionalisation for the better working of internal sovereignty in the state. In the Western world, and particularly in the European Economic Community, agreements by governments on limited devolution of economic sovereignty as the possible precursor to a political union change the perspective on problem areas and revalue the regional character of most frontier zones. The frontier zone tends to display both beneficial and negative aspects of the use of resources, the character of settlement and population structure, arising from its peripheral location relative to the adjacent state areas. The impact of supranational policies in the frontier zone may not be the most dramatic of inter-state associations but it can give rise to considerable friction and, where the frontier zone corresponds to an existing economic problem area, can carry important implications for the regional development policies of the participant states.

THE FRONTIER ZONE

While it may be accepted that the frontier is by tradition a zone, whereas the political boundary is a delimited and usually demarcated line within that zone, it is surprising that so little attention seems to have been given to the outer delimitation of the frontier zone within states on either side of an international boundary.

Lösch's work on the financial sphere of El Paso, on the United States–Mexican boundary, implied the definition of hinterlands extending variably from some point facility, in this case the banks in El Paso, variably into the territory of both adjacent states. This is a sophisticated element of service industry but indicative of the very great variety and complexity of the problem of ascertaining acceptable inner limits of the frontier zone definable often by a multiplicity of criteria. In essence one is initially investigating the zone structure of the diminishing influence of the boundary line, in economic and social terms, as one moves away from it into the area of adjacent states. It may, of course, be argued with some justification that only in a limited sense does movement away from the boundary diminish its significance. The realities of inter-state contact may most severely affect particular industries or communities in any part of the state territory; alternatively, for example, all the farmers throughout a state may be directly affected by some change in economic foreign policy. Nevertheless elements of the frontier zone can be, and traditionally have been, given more precise territorial expression, because the proximity of the boundary has had a detectable influence on the lives and adjustments in border communities.

In a negative sense economic and social life along militarised frontiers is disrupted, and the use of the land may be limited or indeed totally sterilised, either temporarily or permanently. Such restrictions on access to resources were commonplace throughout inter-war Europe but are less common – though not unknown – today along sections of the Iron Curtain. At its narrowest territorial definition, the problems of life near a political boundary are most sharply expressed in the band of communities (local authorities) immediately flanking the line of sovereignty. These may have become adjusted to the existence of the boundary but the latter nevertheless remains a dominant, and often inhibiting factor in local living conditions. If the frontier zone has poor resources or is inaccessible, the problems may be magnified by the political factor, and issues such as rural depopulation, deterioration of social capital or inadequacy of infra-structure are more strongly highlighted.

The difficulties of living close to an international boundary line

have at various times in the past been recognised by the grant of certain privileges relating to local trading or trans-border movement of people within a legally defined zone acceptable to both flanking states. This zone has been variously designated in treaties as the immediately adjacent boundary communes, or quite commonly as a 10-km zone on either side of the boundary, for example in the Franco-Italian Convention of 1951. Occasionally, as in the Italian Peace Treaty of 1947, a demilitarized zone of 20 km was further decreed on the Italian side of the new French boundary. Along the Italo-Yugoslav boundary in recent years a zone of 15 km, later extended to 30 km, was designated a zone of special customs privileges, and in some respects restriction, for border dwellers.

Against the economic disadvantages likely to accrue from living close to an international boundary, certain advantages may occasionally occur, though they are usually much more limited in character and impact. At or near the major crossing-points there is often aggregation of special services at establishments located as close to the boundary as possible in order to take advantage of distinct price differentials between states. Tourist shops, hotels and restaurants are characteristic examples, but in some cases, as on the United States–Canadian boundary through the Great Lakes, or in south-east Limburg between the Netherlands and West Germany, industrial plants also tend to congregate just within a tariff wall, though they may draw their workers from both sides of it. Alternatively, labour cost differentials between states may themselves attract branch factories to locate within journey-to-work distance of nationals in an adjacent state. The price differentials on consumer goods may also strengthen the growth of an exotic border economy confined to settlements within at most a few miles of a boundary line and not sustained by any sizeable hinterland within the domestic territory.

The positive stimulation to the economy of border settlements and the negative inhibiting effects of the boundary line, the latter regrettably hitherto much more common, usually occur simultaneously at different points within the same frontier zone. Lösch drew attention to the frequent coincidence of depressed areas with frontier districts, a state of affairs apparently most common in urbanised industrial countries with their sharp gradient in living standards between the towns and the rural perimeter.

BOUNDARIES AND FRONTIERS WITHIN THE EUROPEAN ECONOMIC COMMUNITY

The long-term implications of implementing the Treaty of Rome may be far-reaching for the internal frontier zones within the European Economic Community. The intention of proceeding by limited stages in devolution of sovereignty, through an initial Customs Union towards closer political association, if not ultimately federal or unitary statehood, is an important prototype for many Common Market systems currently being envisaged elsewhere. Since many of the political boundaries between the European Economic Community countries are of long standing, antedating the major industrial and urban developments of the nineteenth century, some initial dislocation in the frontier zones may be expected and much reorientation of contacts is to be envisaged during the rapid link-up of member states which is now in progress. The great interest for the political geographer here lies in reviewing, almost as a controlled social and economic experiment, the changing significance of political boundaries within the Community during its most formative period in the next decade or so.

The frontier zones and types of boundary feature in the European Economic Community, many of them deeply embodied historically, are remarkably diverse for such a limited area. In the south the Franco-Italian boundary follows the watershed of the western Alps, through high mountain terrain, whose economy is by tradition sylvo-pastoral, with development of hydro-electric power and some localisation of high-energy consuming industries in the valley floors of Savoy and the Val d'Aosta of northern Italy. This boundary has fluctuated during historic times, and at one stage embraced the territories of the Dukedom of Savoy, astride the Alpine watershed. Since the Treaty of Utrecht (1713) it has been in principle a watershed boundary, with a departure in the Maritime Alps, where the frontier lay at intervals along the lower Var, notably from 1815–60. By the Treaty of Turin (1860) the Franco-Italian boundary was moved up to the French glacis of the Alpine watershed, and in 1947 to the watershed line itself. In the 1947 Treaty modifications were made at the principal cols, moving the boundary on to the Italian side of the watershed. Except at the cols the contacts between dwellers on opposite flanks

of the mountains have been limited to the high pastures and certain communal woodlands.

The Franco-German boundary in Alsace-Lorraine has had a fluctuating movement over the past century, and indeed throughout the history of the marchland between Teuton and Frank. The successive consolidation of French sovereignty towards the strategic frontier of the Rhine was followed by the 1870 boundary which cut across the iron-ore field of Lorraine and separated the textile industry of the western Vosges from that of Alsace. Restoration to the line of the Rhine in 1919 and 1945 emphasised the economically integrating role of the river basin in its upper course. The Palatinate and Saar boundaries followed the meandering limits of historical divisions, cutting across the Saar–Lorraine coal basin, while farther west the Luxemburg–Lorraine boundary cut through the iron-ore field. The run of these boundaries powerfully influenced the settlement and pattern of economic development on either side of the boundary. In the former Saar territory and Luxemburg a fascinating possibility exists for the comparative study of two small areas of sovereignty in a buffer-state location.

The northern France–Flanders boundary today cuts through the coalfield and industrial region; at its demarcation, in 1713, only the coal reserves of the Low Countries were known and these were purposely kept within one state. The Ardennes frontier follows an old forest boundary with a similar rural economy on both flanks.

The Low Countries' boundaries date from the Treaties of Paris and Vienna (1815), though the internal boundary, following the partition between the Netherlands and Belgium, dates from the Treaty of London (1839). The latter thus created the appendix of South Limburg, economically a part of the Zollverein until 1867. Only minor boundary modifications have since occurred; Eupen–Malmédy and small tracts passed to Belgium in 1919 and some minor modifications were made to the Dutch–German boundary after 1945.

Though the Ardennes frontier is sparsely settled and has an extensive type of sylvo-pastoral economy, the situation changes sharply immediately to the north, where the Liège–Maastricht–Aachen–Ruhr axis forms the economic and industrial backbone of the entire Economic Community. A total revaluation of political boundaries is imminent around the appendix of southern Dutch Limburg, now at the crossroads of Western Europe, but until recently focused almost entirely on north–south contacts as an isolated frontier province of the Netherlands.

The Dutch–Belgian frontier across the marshes and heaths of the Peel and Brabant passes through marginal land, with the important Campine coal basin and developing industrial zone to the south separated from manufacturing towns such as Eindhoven to the north. The problem of divided sovereignty on the lower Scheldt is paralleled on a grander scale by the Dutch sovereignty over the lower reaches of the Rhine system, a region with an immense potential for development as the economically inhibiting effects of divided sovereignty in the Rhine basin are progressively abated. In the north-eastern Netherlands the heathlands and moors separate the border provinces from those of West Germany. Post-war oilfields have developed on both sides of the boundary, while the Dutch Development Area policy has created a transformation in the borderlands of Drenthe province, for which there is as yet no counterpart on the German side.

The problems of the internal Economic Community boundaries thus fall into two categories. The first consists of areas of poor resources, often with an isolated location and built-in social and economic problems of stagnation or decline, necessitating priority aid under government policy, for example the central and southern French Alps and the mountain zone of northern Italy, the Palatinate and the Eifel, both border districts undergoing depopulation with a decaying, outdated rural economy, and the heaths and moors of the north-eastern Netherlands. This category is marginal because of frontier location which has aggravated the basic problems arising from environment and history. It is difficult not to expect further decline in such areas, even though limited intergovernment aid may continue to cushion the effects of the developing Economic Community upon such underprivileged areas.

A second category of problems concerns the frontier zone in areas of partitioned natural resources, principally of coal and iron ore, where the boundary produces an impediment to the free flow along the logical routes provided by the river systems or the shortest routes linking major industrial regions. The Saar–Lorraine field, the northern France–Belgium coalfield, the South Limburg–Aachen coal basin, the Lorraine–Luxemburg iron-ore deposits and the petroleum and natural gas fields on the German–Dutch

border come into this category. In most respects the development of the Common Market will foster a more logical use and interchange of resources and stimulate the co-operative expansion of the partitioned coal, ore and petroleum fields. Yet in other respects governments and private enterprise have followed national investment programmes and there has been duplication of facilities on both sides of the boundary. Rationalisation may thus require some closures of plant, while others which were originally located to benefit from the artificial financial advantages of a frontier site will lose their *raison d'être*. Under the Treaty of Rome governments have a moratorium of some years in which to equalise their industrial location policies and reduce aid to those marginal districts, including frontier zones, which cannot be justifiably supported within the wider Common Market.

To focus the issues in the frontier zones of the Economic Community two contrasting case studies will be briefly examined: the Alpes Maritimes (France–Italy), and South Limburg (a contact zone between Netherlands, Belgium and Germany).

THE ALPES MARITIMES

The uplands along this boundary (see Fig. 23.9) have one of the most depressed economies in the Economic Community borderlands. The traditional sylvo-pastoral way of life has been in decay for many decades, a decline accompanied by steady depopulation to the nearby Riviera coast. The gradient in living standards over such a short distance is one of the most remarkable in Western Europe. In the immediate hinterland of the Mediterranean coast the peasant economy, traditionally based on the olive, chestnut and the vine, has also been declining. The added political ingredient of a boundary change in 1947 further deepened the economic and social problems of this frontier zone.

In the frontier zone along the crestline between the headwaters of the Tinée, Vésubie and the Roya in France and the tributaries of the Po system in Italy there is a similar economy on both flanks of the Alps. The transhumant movements with cattle, and occasionally sheep, have ranged widely, and legal rights of both the individual and the commune are tenaciously held. Wealth is locked up in the communal woodland, the stinted pastures and the herds and flocks which follow traditional season-by-season transhumance routes. These did not respect the 1859–60 boundary, though with time they became gradually adjusted to it, nor have they entirely conformed to the crestline of the Alps. The 1947 boundary dislocated the pastoral economy of the communes east of Tende and La Brigue. In 1954, after skilful arbitration by a Swiss pastoral expert, agreement was reached on re-apportionment of woodland and pastures, but structural damage to the rural economy had already taken place. Rural population had been lowered and the density of stock had fallen.

Farther south the international boundary cuts across the lower reaches of the Roya valley, with some southward adjustment in 1947 from the strategically important Mt Grazian area formerly in Italy. In this frontier zone of lower altitude the rural densities of population are higher and the farmers are mainly sedentary. The wealth of the communes lies in their olive and chestnut groves and small vineyards. Water is a scarce resource and virtually controls both the size of rural populations that can be maintained and the productivity of the land. The social hinterland of each village is very limited, and life is more self-contained than in the semi-mobile transhumant society farther north. The rural economy of these border villages is decaying, a decline accelerated by the location of some near to the new international boundary. Within France the Government has developed a rural support programme for improving the living conditions and the infrastructure of the more isolated settlements.

The Lower Roya was partitioned under the Treaty of 1860; in 1947 France added possession of the headwaters to those of the middle reaches, but the lower reaches were retained by Italy. The partitioning of the valley made realisation of its hydroelectric power resources a piecemeal affair. More stations were built than were economically desirable; all were of limited potential, and re-apportionment of the upper valley has further complicated the production and export of electric power.

Apart from the heavy tourist traffic along the Riviera coast, this frontier zone remains remote from the major circulation currents of the European Economic Community, and the Nice–Cuneo railway, destroyed in the war, has never been restored. The contacts between opposite sides of the boundary remain limited and there is common economic distress in the border communes of both France and Italy. This led in 1951 to a Franco-Italian convention

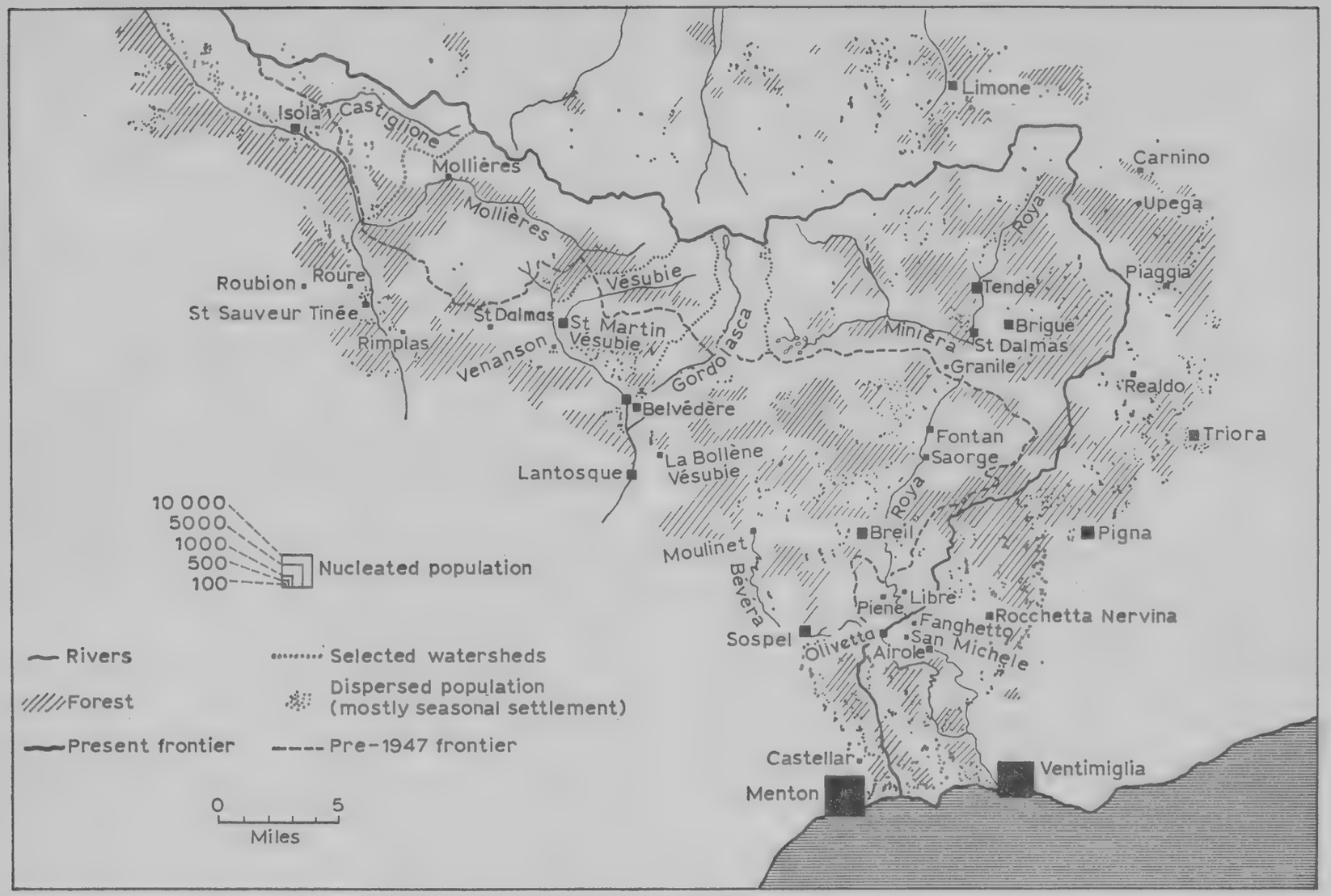


Fig. 23.9 Rural frontier zones in the Alpes Maritimes

defining a frontier zone 10 km in depth from the boundary, within which movement of persons and products was to be stimulated and protected, in the interests of border dwellers. Latterly both governments have been jointly approached by the local inhabitants of the frontier zone, asking for their problems to be given special recognition.

THE LIMBURG APPENDIX

In almost all respects this region (see Fig. 23.10) is the antithesis of the Maritime Alps. Its economic structure is partly urban and industrial, with coal mining, in direct contrast to the rural economy of the Alps. Its location, immediately north of the main west–east European axis from northern France and Belgium to the Ruhr, close to where the latter intersects the route from the Randstad cities to southern Belgium and eastern France, is nodal, not peripheral to the economic heart of Europe. Hitherto the advantages of such nodality have been masked by the divided sovereignty of the area, with South Limburg forming an ill-shaped and distant appendage of the Dutch state, remote from the populous urban areas of the lower Rhine and Holland province. Indeed with a rapidly developing population and limited economic resources the area has become a Development Area under Dutch economic planning legislation. The traditional trade linkage has been north to south through Limburg, with little consideration for its potential in respect of west–east movement. Even the Benelux agreements after 1945 conferred little direct benefit upon Limburg.

The frontier zone covers the width of the appendix, which at its narrowest is only three miles, and extends into the Campine to the west, the South Limburg loess-covered plateau and Pays d'Herve to the south, and the Aachen coal basin to the east. In contrast to the zone in the Maritime Alps, there is here a lively interconnection of border traffic passing in both directions across the Dutch frontier. Unemployment in Dutch Limburg is high and labour costs are lower than in either Germany or Belgium. Thus there is a well-developed journey-to-work traffic, into both West Germany and Belgium, including some seasonal and semi-permanent labour migration. Belgian workers cross into the frontier town of Maastricht for daily work, even though the direct wages are lower;

Fig. 23.10 The Limburg appendix

work is lacking in adjacent areas of Belgium but a compensating system of salary allowances has been introduced.

TABLE 23.1

SOUTH LIMBURG: INTERNATIONAL JOURNEY TO WORK, 1959

Limburg–Belgium	2600
Limburg–Western Germany	2100
Belgium–Limburg	3600
West Germany–Limburg	700
Total	9000 workers per day

The communications pattern has dominated the development of South Limburg, which for heavy traffic is focused on the Maas waterway. The competition between Antwerp and Rotterdam for the upper Maas hinterland led between the wars to the construction of national canals, namely the Albert Canal, by Belgium, from Liège to Antwerp, and the Juliana Canal, by the Netherlands, from Maastricht to the lower Maas. Both were capable of taking 2000-ton barges but were not interconnected; the Ternaien lock on the Liège–Maastricht Canal could not take craft of more than 600 tons and until 1936 the limit was only 400 tons. National interests kept this bottleneck in being until 1963, when a new lock and canal section linked the Albert and Juliana canals, through Ternaien, thus opening interconnected routeways for 2000-ton barges from the Rhine–Ruhr to the upper Maas and the lower Scheldt.

The knot of European motorways has focused immediately north of Liège, but the Limburg appendix will soon be more closely in touch with the new feeder network of these roads. Its position is capable of being transformed from that of a remote outer province of the Netherlands to a nodal industrial area astride the major European routes, a transformation which should logically follow the elimination of the divisive political factor of the boundary lines around the appendix.

The effects of the development of the European Economic Community on the frontier zone of Limburg will be more complex than these changes in the communications pattern alone might suggest. In the first place the industrialisation policy successfully followed by the Dutch Government has already tied the province more closely to other parts of the Netherlands. Fortunately it has been a policy of stimulation rather than subsidised inducement to industrialists, a development based on growth points rather than depressed areas. Reorientation within the wider Common Market economy will thus be facilitated, though it is interesting to note that fifteen years of the Benelux Union surprisingly conferred little benefit on Dutch Limburg.

There have been exotic growths within the frontier zone, inspired by the price differentials in favour of the Netherlands and against her neighbours; foodstuffs, particularly butter and meat, are much cheaper in Dutch Limburg. The frontier town of Vaals, almost a suburb of Aachen in the extreme south-eastern corner, demonstrates this point clearly. The Dutch hinterland of the town is very restricted, but its diversity and number of shops and banks establish the town as an important and thriving service centre. An increasing proportion of its trade is across the border for shoppers from West Germany, who benefit from the lower Dutch prices. Branch plants of the Aachen textile industry have also been set up in the town to utilise more abundant and cheaper Dutch labour and to serve the Dutch market in part from within its national boundaries.

The equalisation policies of the Economic Community will diminish the price differentials which at present work in favour of the service settlements within Dutch Limburg and against the German border town of Aachen in particular; likewise there will be less incentive for Dutch workers to go abroad for higher wage-levels. The Limburg farmers should benefit from higher prices in the Community, whose major urban centres are more accessible than those of the Netherlands.

Unlike the Western Alps, Dutch Limburg is poised to take advantage of a revaluation of its European location, freed from the hitherto limiting effects of its national frontiers. At the same time it will need to be more closely linked with the Liège–Aachen axis, already firmly established to the south of formerly neutral Dutch territory.

CONCLUSION

These two examples show almost the extremes within the spectrum of frontier zones in the European Economic Community. Even during the next decade the tempo of economic freedom across

political boundaries will sharply reduce their traditionally entrenched nature, and closer contacts, through the greater exchange of commodities and the freer movement of workers, will develop within the equalisation policies envisaged by the Treaty of Rome. The revaluation of boundaries will have distinctive effects upon those living within each of the varied frontier zones. Some of these zones have economic and social problems aggravated by the impediments to circulation which political boundaries have by tradition imposed. In others, settlements have been artificially stimulated by the development of land and building use to profit from the inter-state economic differentials at individual crossing-points. Both types may exist side by side in the same frontier zone, and both may need a measure of government protection in an interim period of readjustment. Such a policy would be a part of the regional programmes to which most governments in the Western world are now committed. What is worked out for the frontier zones of the European Economic Community in the next decades may later be of interest in other parts of the world as the devolution of absolute sovereignty spreads, and the once barrier-like qualities of political boundaries are progressively reduced in the interests of freer and fuller international trade.

The two pieces which follow have been extracted from books which are intended to introduce students to aspects of government and politics. The first extract is taken from Robert Dahl's book 'Modern Political Analysis'. In this extract, Dahl deals principally with the question of the classification of 'political systems' and introduces two classical definitions – Aristotle's and Weber's. He also examines the criteria from which classification is made. The second extract, 'Constitutionalism: A Preliminary Discusssion', is by G. Sartori, and deals in a clear, straightforward way with the question: what is a constitution?

POLITICAL SYSTEMS: DIFFERENCES

Robert A. Dahl

THE MULTIPLICITY OF CRITERIA

THE differences among political systems are more important to most of us than their similarities. It is the real or presumed differences in political systems that make one precious and another hateful. Human beings have to live in political systems; but they do not have to admire them equally. How, then, can we distinguish one political system from another? Obviously there are thousands of different possible criteria. Out of all these, which criteria we find most useful depends on the aspects of politics in which we are most interested. A geographer might distinguish political systems according to the area they occupy, a demographer by the number of persons who are members, a lawyer according to their legal codes. A philosopher or theologian interested in distinguishing 'the best' political system will use ethical or religious criteria. A social scientist interested in determining how revolution is related to economic conditions might classify political systems by relative income and frequency of revolutions. Just as there is no best way of classifying people, so no single way of distinguishing and classifying political systems is superior to others for all purposes.

TWO FAMOUS SCHEMES

Aristotle's classification

Aristotle's classification has, however, proved so useful for so many different purposes that it has survived the political vicissitudes and transformations of twenty-five centuries. Indeed, it or something like it is as much a part of the habitual thinking of the student of politics as it was to the observant Greek in Aristotle's day.

Aristotle proposed two criteria for distinguishing states or 'constitutions'. (These criteria can be applied not only to states but to political systems in general.) The criteria were 'the nature of the end for which the State exists', and 'the various kinds of authority to which men and their associations are subject'.[1]

Aristotle used the first criterion to distinguish between systems in which the rulers govern in 'the common interest', in which, that is, they seek the 'good life' not simply for themselves but for everyone in the system; and systems in which the rulers pursue their own selfish interests rather than 'the common interest'. The first he regarded as 'right' constitutions, the second as 'wrong' constitutions, or perversions of the first type. Aristotle used the second criterion to distinguish systems according to the relative number of citizens who were entitled to rule; in this way he arrived at the familiar distinction between the rule of One, Few or Many.

> On this basis we may say that when the One, or the Few, or the Many, rule with a view to the common interest, the constitutions under which they rule must necessarily be right constitutions. On the other hand the constitutions directed to the personal interest of the One, or the Few, or the Many, must necessarily be perversions.

Thus Aristotle produced a famous sixfold classification illustrated in the table opposite.

Two aspects of this scheme are particularly interesting. First, Aristotle combined both *empirical* and *normative* criteria in his classification; that is, his classification is not intended merely to describe what *is* but also to instruct us about what *ought* to be. To

		Rulers rule in the interests of:	
		All	*Themselves*
The number of citizens entitled to rule is:	*One*	Kingship (Monarchy)	Tyranny
	Few	Aristocracy	Oligarchy
	Many	'Polity'	Democracy

classify a political system according to the criterion of number is, in principle, a matter of observation. Whether one person, a few or many rule in a specific political system is a question of fact (verifying what *exists*) and not a question of value (determining what *ought* to exist). On the other hand, to classify a system according to the second criterion involves a mixture of both fact and value. One must not only observe the actions of rulers in order to learn what the aims or consequences of their actions *are* but also appraise their actions according to some criterion of 'interest' or 'rightness' so that one could say whether or not these actions are what they ought to be. To judge whether rulers rule in their own interests or those of 'all', then, requires standards of value.

In the second place, Aristotle gave the name 'democracy' to the perverted form of popular rule, whereas in his day as in ours adherents of democracy would insist that what they mean by 'democracy' is the rule of Many in the interests of all. Aristotle, however, held a slightly jaundiced view of Athenian democracy; like many later conservatives, he feared that the rule of the Many would typically lead to the tyranny of a poor and propertyless majority over the middle classes. As often happens in political discussions, Aristotle may have intended to discredit the term 'democracy' by defining it as a perverse rather than a good form of government.

Curiously enough, Aristotle was dissatisfied with his own classification scheme. For he assumed it to be a fact that the Many would usually be poor and the rich would be few. Consequently (he went on to argue) the criterion of number was unsatisfactory; the essential characteristic was really the ownership of property:

> The real ground of the difference between oligarchy and democracy is poverty and riches. It is inevitable that any constitution should be an oligarchy if the rulers under it are rulers in virtue of riches, whether they are few or many; and it is equally inevitable that a constitution under which the poor rule should be a democracy.

In introducing differences in property or social class as a criterion for distinguishing political systems, Aristotle muddled his original classification scheme. Yet the price may have been worth it, for he was calling attention, though in a somewhat unclear manner, to a highly important hypothesis: namely, that the amount of power and influence members of a political system have depends on how much income, wealth and social status they have. To the extent that income, wealth and status are distributed differently in different systems, then we should expect differences in the way power is distributed in each system. Thus Aristotle sought not only to *classify* political systems according to certain key differences but also to *explain* how these differences arose.

Weber's classification

Where Aristotle was concerned with the numbers and social positions of rulers and the character of their authority, Max Weber devoted himself almost exclusively to the nature of the authority held by the rulers. In fact, as mentioned earlier, Weber restricted his attention to systems in which the government was accepted as legitimate.

Weber suggested that the leaders of a political system might claim legitimacy for their rule, and members might accept their claims, on three grounds:[2]

1. *Tradition.* In this case legitimacy rests 'on an established belief in the sanctity of immemorial traditions' and in the need to obey leaders who exercise authority according to the traditions. Weber held that this was 'the most universal and primitive case' of authority. But even modern systems acquire a good deal of legitimacy from their traditions. In few countries has political tradition had more influence than in England. Many aspects of modern British politics, such as the monarchy, are accepted by British subjects today because of immemorial tradition – after all 'England has always had a monarch'. Probably every stable political system gains some measure of legitimacy through the workings of tradition. Even in the United States, where people sometimes take pride in their freedom from tradition, political practices are often defended on purely traditional grounds. The right of the Supreme

Court to declare acts of Congress unconstitutional cannot be found anywhere in the Constitution. Yet few Americans propose to take this right away, and doubtless many would defend the institution by saying that 'this is the way it has always been'.

2. *Exceptional personal qualities.* In this case legitimacy is based on 'devotion to the specific and exceptional sanctity, heroism or exemplary character of an individual person' and the moral or political order he has revealed or ordained. This is the kind of legitimacy the inspirational leader gains for his leadership – Churchill, Roosevelt, de Gaulle, Lenin, Hitler, Huey Long, Mahatma Gandhi, Fidel Castro. As the list of names suggests, an inspiring leader may be a demagogue or a statesman; but he has an unusual capacity for transmitting to his followers a sense that he pursues grandiose goals, more lofty than mere self-interest, and is worthy of admiration, awe, imitation and obedience.

3. *Legality.* In this case legitimacy rests on a belief that power is wielded in a way that is legal; the constitutional rules, the laws and the powers of officials are accepted as binding because they are legal; what is done legally is regarded as legitimate. Weber believed that legality was the usual basis of legitimacy in the modern era, but he might have revised his opinion if he had lived to witness the rise of revolutionary and nationalist leaders in Europe, Asia, Africa and Latin America. None the less, in much of the modern world and in many different kinds of organisations, it is true that laws, rules and constitutional practices acquire their legitimacy in large part because the process by which they are enacted is assumed to be legal. Whenever legality is the basis for legitimacy, leaders who are known to act illegally are discredited and lose their authority.[3]

To each of these three main grounds for legitimacy, then, there corresponds a 'pure' form of authority; (1) traditional authority, (2) 'charismatic' authority, from a Greek word used by early Christians meaning 'the gift of grace' and (3) legal authority.[3]

Weber recognised that these 'pure' forms were abstractions or, as he called them, 'ideal types'. In an actual political system one might encounter all three kinds of legitimate authority – as, indeed, our choice of examples has already suggested would be true in the American national government.

Like Aristotle, Weber also sought to show how one kind of authority might change into another. The Presidency of Franklin D. Roosevelt can be used to illustrate Weber's ideas. The traditional authority attached to the Presidency was expanded by Franklin D. Roosevelt who had an immense popular following; Weber would have called Roosevelt a 'charismatic' leader. Then, when Roosevelt's reform proposals were enacted into law, 'bureaucracies' (as Weber would call them) were created to administer the new laws – as in the case of the National Labour Relations Act, the Social Security Act and the like. Thus Roosevelt's *charismatic* authority enabled him to break through the *traditional* limits on the authority of the Chief Executive, which, in turn, was transformed into *legal* authority. Weber believed that charismatic authority was inherently unstable, and tended to be transformed by subordinates and successors to the inspirational leader into either traditional or legal authority. Both these forms of authority, he held, were more enduring, since they did not depend on the existence of a particular leader who possessed 'charisma'.

There are obviously some important differences between Aristotle's scheme and Weber's. For one thing Weber dealt only with authority – that is, with legitimate power. His scheme does not concern itself at all with illegitimate regimes. Weber justified his concentration on legitimate governments on the ground that every system of authority 'attempts to establish and to cultivate belief in its legitimacy. . . .'

In the second place, unlike Aristotle, Weber did not distinguish political systems according to their moral value. To Weber, even legitimacy was a purely empirical concept; the test of legitimacy was what the members of the political system thought of their leaders, not what Weber or someone else thought. The fact that a leader's power was legitimate did not, to Weber, make that system good.

Curiously enough, Weber has been criticised both for omitting standards of evaluation and for including them. Some critics have interpreted him as saying that all legitimate governments are good – and equally good. Yet on any scrupulous reading of Weber it seems clear that he harboured no such attitude. He intended his categories to be neutral, descriptive and scientific rather than ethical or 'normative'. Unlike Aristotle, he sought to exclude any criterion that would depend either on his own moral evaluations of different forms of authority or on an appeal to specific standards of value that he might expect his reader to share. Some critics

have attacked Weber precisely because he did seek to be 'value-free' and 'scientific'. This dispute over the proper or inherent place of values in such political analysis has received a great deal of attention from political theorists in recent years.

Even as a scientific classification, however, Weber's typology seems deficient, since it makes no place for a number of distinctions that most students of politics would regard as interesting and significant. For example, his scheme does not distinguish a stable democracy from a stable monarchy, for both *could* contain about the same mixture of traditional, charismatic and legal authority. In Weber's defence, it might be said that he was less concerned with political systems than with the various types of authority, and 'pure' types at that. But some significant kinds of authority also seem to be missing from Weber's scheme, or else several different forms of authority are confounded under the same heading. Where, for example, would one locate the authority of the expert? An expert's authority is not necessarily grounded on an appeal to legality, and ordinarily it is neither traditional nor charismatic.

CONCRETE SYSTEMS AND ABSTRACT CLASSIFICATIONS

The ways of distinguishing political systems proposed by Aristotle and Weber both have drawbacks as well as advantages. As we stressed at the outset, this is likely to be true of any system of classification. To repeat what was said at the beginning of this chapter, there is no one best classification scheme for political systems. How useful any given scheme is depends on one's purposes.

A classification is a way of simplifying – and, in this sense, 'distorting' reality. All empirical analysis requires some simplifying 'distortion'. If every atom were treated as unique, physics would be impossible; if every medical case were considered unique, no advances in medicine could occur. In empirical analysis knowledge consists of generalisation, which requires one to exclude the unique in favour of the general properties of a concrete event or system.

In the real world, concrete systems rarely represent 'pure' types; they are usually *mixed*. This is true in physics, biology, medicine and, of course, politics. No matter what scheme of political classification one might propose, it would be difficult to find concrete systems that exhibited only the characteristics used in that scheme. Consider any national political system, that of the United States, for example. In some respects the American system is democratic. We have frequent elections, widespread suffrage, competing political parties, peaceful assumption of office by winning candidates and so on. But some aspects of the system, such as the postal department or the military, are hierarchical, or, one might say, monocratic. In the relations among leaders in the executive branch, Congress, the political parties and private organisations, there is also an element of bargaining. Consequently, when one classifies the American political system as a democratic system, he necessarily abstracts one aspect of the system out of the concrete reality. It could also be classified as a bargaining system or, for certain purposes, even as a hierarchy.

Despite these inherent difficulties, in any method of classifying concrete objects some criteria are usually related more closely than others to a wide range of phenomena. In politics, three differences among political systems are closely related to a wide variety of other aspects of a system. If systems differ in one or more of these ways, they are likely to differ in a great many other ways as well. These three important differences are:

1. The extent to which the government is legitimate and the power of leaders is accepted as authority.
2. The proportion of the members who influence the decisions of the government.
3. The number of sub-systems and the amount of independence they have.

Notes

1. E. Barker (ed.), *The Politics of Aristotle* (Oxford, 1948).
2. M. Weber, *The Theory of Social and Economic Organisation*, trans. A. M. Henderson and Talcott Parsons (New York, 1947).
3. A puzzling feature of Weber's general theoretical scheme arises from the profusion of classification schemes in his unfinished work *Wirtschaft und Gesellschaft*. These schemes are based on the application of essentially the same criteria to a number of different aspects of the social system; sometimes, however, there appear to be four main criteria and sometimes only three. Evidently, by the time he came to discuss the basis of legitimacy in the final sections of his volume, he had dismissed the fourth of these; had he employed it, one might have expected an analysis of a system of authority based on the acceptance of an absolute standard of value.

CONSTITUTIONALISM: A PRELIMINARY DISCUSSION

Giovanni Sartori

In the nineteenth century what was meant by the term 'constitution' was reasonably definite and clear. Paradoxically enough, if the word retained some ambiguity, this was because of the British constitution; that is, because the mother country of modern constitutionalism appeared to have an obscure constitution, or even – according to some of the standards that seemed very important elsewhere – no constitution at all.

I

Yet the very term 'constitution' has acquired its modern meaning in English, in the course of the evolution of the English legal terminology. The Latin term *constitutio* meant the very opposite of what is now understood by 'constitution'.[1] A *constitutio* was an enactment; later, after the second century, the plural form *constitutiones* came to mean a collection of laws enacted by the Sovereign; and subsequently the Church, too, adopted the term for canonical law. The terms *constitutio* and *constitutiones* were not frequently used, however, by the English medieval glossators (while frequently used, as a synonym for *lex* and *edictum*, by the Italian ones). This explains why, in the course of time, the word constitution became a 'vacant term' – i.e. a term available for a new employment – in English (this does not necessarily mean in England), and not in those languages which had retained the Roman legal terminology.

For terminological reasons also, then, we have to refer to an English meaning of the word constitution, although the United Kingdom has a difficult and *sui generis* constitution, deriving from a tortuous sedimentation of common law, acts and conventional usage, partly legal and partly extra-legal, and despite the fact that, when one reads the British constitutional lawyers, one is often reminded of what was said in a review of Stirling's book, *The Secret of Hegel:* 'never has a secret been better kept'.

For one thing, English constitutional lawyers appear to take a particular pleasure in pointing out to foreign jurists and political thinkers (beginning with Montesquieu) that their understanding of the English system is quite wrong. To be sure, this has been and still is very often the case. But one remains with a feeling that the British find a special gratification in confounding alien scholars: there is an element of polemic coquetry in the emphasis they lay on the principle of the supremacy of Parliament (exhibited as being unlimited, arbitrary, omnipotent, *supra* and *contra legem*, etc.); in the somewhat provocative and bold statement that, according to the American and French meaning of the term, the United Kingdom does not have a constitution;[2] in the point that the British system is based not on the 'division' but on the 'fusion' of powers; or in the way in which Sir Ivor Jennings puts forward that, 'Since Great Britain has no written constitution, there is no special protection for "fundamental rights".'[3] And one could quote at length.

All these statements are, to be sure, true. But they are 'literally true', and one is brought to wonder why the emphasis is laid on the *letter* so much more than on the *spirit* of the law of the constitution. After all, constitutionalism has a prescriptive purpose; whereas English scholars appear more inclined to address themselves to an M.P. by saying 'you could' rather than 'you cannot'.

Let us take, for instance, the principle of the supremacy of Parliament. Would it be far from the mark to say that if the principle is related to the historical circumstances of its establishment, it hardly carries with it the dangerous implications that British scholars somewhat proudly expound?[4] Parliament, in the English terminology, means the King, the Lords and the Commons acting together as the supreme governing body of the realm. Thus, if the principle of the supremacy of Parliament is translated into Continental terminology, it amounts to what is otherwise called the 'sovereignty of the State'. Historically speaking, moreover, the principle of the supremacy of Parliament is the counterpart of the principle of the supremacy of the Crown, and what it really meant when it was affirmed was that the King had no power *outside* of Parliament, that his prerogatives could only be exercised according to the formula of the King *in* Parliament. If this be so, would it be very wrong to conclude, despite the contrary opinion of British scholars, that parliamentary sovereignty in England actually contradicts the idea of a 'higher law' no more than any flexible constitution does, and that the conventions of the constitu-

tion hardly allow a parliamentary majority to pass *any* law whatever?[5] (The point being the 'any'. Let us remember that the difference between accepted political behaviour and wicked Machiavellian politics is rendered by the difference between the sentence, 'The end justifies the means', and 'The end justifies *any* means'.

Probably I shall be told that I am wrong. However, as I was trying recently to sort out the juridical features of present-day dictatorial systems, it occurred to me that some English constitutional textbooks would be of great help. 'No higher law and limitless legislative power', 'fusion of powers', 'no special protection for fundamental rights', etc. – all these principles applied very nicely to dictatorships. What is more, the very definition given in Wheare's textbook. 'The British constitution is the collection of legal rules and non-legal rules which govern the government in Britain', appeared to be – but for the terms 'British' and 'Britain' – the most fitted to describe dictatorial constitutions (For 'rules' applies to any kind of rule – including the rule of rulership – and the specification 'non-legal rules' just gives the final touch of perfection to the concept of unchecked, uncontrolled and absolute exercise of power.) I have not made the experiment the other way around, that is, using the apologetic writings on dictatorial constitutions in order to get highlights on the virtues of the English one: but my guess is that, again changing but one word (the geographic referent), the English constitution would come out in brighter colours following this procedure than it appears, e.g. to the reader of Jennings's *The Law and the Constitution.*[6]

So, I may well be wrong. But something must be wrong too in the way in which many English scholars understate their constitution, seem to make a particular point of not being helpful and leave the alien reader with the feeling that the British constitution really amounts to the fact that, in the final analysis, the British people are clever and fine people who know how to go about in politics. I am personally convinced of this. But allow me to repeat that this conclusion is not very helpful.

I have drawn attention to the British habit (and perhaps coquetry) of understatement, for unless this element is taken into account, one is likely to miss some important points. To begin with, this point: that, despite the English 'mystery', in the nineteenth century, all over Europe as well as in the United States, a general agreement prevailed as to the basic meaning of the term 'constitution'. In 1830, and especially during the 1848 revolutions, it was very clear on both sides of the Channel what the people were asking for when they claimed a constitution.[7] If, in England, 'constitution' meant the system of British liberties, *mutatis mutandis* the Europeans wanted exactly the same thing: a system of protected freedom for the individual, which – according to the American usage of the English vocabulary – they called a 'constitutional system'. Having to start from naught, people on the Continent (as was first achieved by the Americans) wanted a written document, a charter, which would firmly establish the overall supreme law of the land. The British too, however, had, from time to time, relied on particularly solemn written documents: the Magna Charta, the Confirmation Acts, the 1610–28 Petition of Rights, the Habeas Corpus Act of 1679, the Bill of Rights, the Act of Settlement, etc. The circumstance that these British 'supreme laws' are not collected in a single document does not really mean that England has an unwritten constitution. I would rather say that the English do not have a codified constitution, i.e. that Britain has a constitution which is written only in part (or, even better, unwritten to a much greater extent than 'written' constitutions are), in a piecemeal fashion, and scattered in a variety of sources.

However that may be, and if it pleases the British to emphasise the fact that they have a constitution which is not written, this question is of secondary importance. I mean that the written, complete document is only a means. What really matters is the end, the *telos*. And the purpose, the *telos*, of English, American and European constitutionalism was, from the outset, identical. If the English vocabulary had not to this day refused to import the word (another paradox!), this common purpose could be expressed and synthesised by just one word: the French (and Italian) term *garantisme.*[8] In other terms, all over the Western area people requested, or cherished, 'the constitution', because this term meant to them a fundamental law, or a fundamental set of principles, and a correlative institutional arrangement, which would restrict arbitrary power and ensure a 'limited government'. And during the whole of the nineteenth century and until World War I, 'constitutions' remained, in the United States, in England and in Europe, different means (technically speaking) which had

nevertheless the same common purpose in view. That is, for almost 150 years 'constitution' has been – on the whole – an unambiguous term.

II

In the twentieth century, in the few decades following World War I, this situation of overall basic agreement has come to an end rapidly and radically. So rapidly and so radically that one must wonder why. The main reason (or, we might say, the specific agent of change) has been, I believe, the following: that legal terminology – to the extent that it affects what Rousseau would have called *droit politique*, political right – shares the same destiny as political terminology in general: that is, it tends to be abused and corrupted. And this is all the more the case in a time in which politicians have become ever more conscious of the 'power of words'.

In our minds, constitution is a 'good word'. It has favourable emotive properties, like freedom, justice or democracy. Therefore, the word is retained, or adopted, even when the association between the utterance 'constitution' and the behavioural response that it elicits (e.g. 'The constitution must be praised, for it protects my liberties') becomes entirely baseless. More precisely, the political exploitation and manipulation of language takes advantage of the fact that the emotive properties of a word survive – at times for a surprisingly long time – despite the fact that what the word denotes, i.e. the 'thing', comes to be a completely different thing.[9]

To be sure, the agent of change has not only been insincerity and the political debasement of language. For technical reasons of their own, jurists too have been gradually covering up the *garantiste* feature of 'constitution'. This technico-juridical explanation cannot be pushed too far, however. It is quite true that Continental constitutional lawyers have never been at ease when confronted with the problem of putting forward the 'strictly juridical' point of view. On the other hand, the fact remains that, more or less until the 1920s, they managed to combine the requirements of 'pure law' with the *telos* of constitutionalism. If one reads, for instance, the *Leçons* of a representative figure of the pre-1848 period, Pellegrino Rossi,[10] one finds both the statement that 'every State has a constitution' (*ex hypothesis*), and the statement that 'a constitution is the law of the free countries'. Let us face it: the two statements are not consistent. Nevertheless they introduce and represent an approach which received wide acceptance in the European juridical literature for almost a century. It is interesting to realise, therefore, why the Continental constitutional lawyers took this approach, and how they found a way of giving consistency to an inconsistency.

Descriptively speaking, and having reference, e.g. to the terminology of the Constitution of Pennsylvania of 28 September 1776, a constitution contains two basic elements: a 'plan (or frame) of government', and a 'bill of rights'. For the framers of the eighteenth-century charters it was self-evident that the two component parts could not be separated: both were needed for a constitution to be a constitution. They did not mean in the least that *any* plan of government amounted to a constitution: they meant that this was the case only when a frame of government provided for a bill of rights and the institutional devices that would secure its observance. Continental jurists, however, were anxious to put their rationalistically trained juridical consciences at ease by finding a 'universal' definition of constitution. And for this purpose they found it expedient to separate the universal trait (the 'plan of government' meaning) from the *garantiste* component. Therefore, often enough, they did come to say – *qua* pure jurists – that any 'frame of government' amounted to a constitution. They said it, but – let me stress this point – they immediately denied it. For they went on to say that 'it had become customary'[11] to use the term constitution in a more specific *garantiste* sense; and therefore that, according to this practice, it was improper to hold that every state was a constitutional state. *Every* state had a 'constitution', but only *some* states were 'constitutional'.

Unquestionably, the distinction is very thin; and moreover it is hard to explain why the phrase, 'It has become customary', should have been accepted by jurists as a juridical argument. The interesting fact is therefore that they did accept it. And they accepted it – among other reasons – because they fully realised that their 'universal' definition had no denotative value: it indicated *everything*, not *something*. Thus, generation after generation, European public law went ahead riding simultaneously two horses: the constitution as any 'State order', and constitutionalism as a specific 'content' of guarantees.

It was an uneasy equilibrium, I grant. It made them vulnerable, exposed. But one can have an Achilles heel and nevertheless survive. So, despite this weakness, I believe my former assertions still hold good, namely: (i) that for almost 150 years 'constitution' has been associated with *garantisme;* and (ii) that the primary agent of change cannot be located in the inner logic of development of the European juristic tradition. Continental constitutional lawyers were hardly in a position to resist change; but they did not start the change. What started the new trend was the impact of the political atmosphere of the 1920s. For a new look about politics was taking shape in those years. 'Feeble politics', so to speak, was giving way to 'intense politics'; that is, the peaceful-legalitarian approach to political relationships was giving way to a warlike view of politics.

I have said 'political atmosphere'. It is fair to add that in some countries it was not only a question of atmosphere. In Italy, and in the 1930s in Germany, jurists were somehow compelled to adopt a merely formal, 'organisational', neutral definition of constitution. This is not surprising, and Italian and German constitutional lawyers can hardly be blamed for this. What is surprising is that in the meantime the British, too, had come to adopt very much the same position.

Let me refer again to Wheare's definition, according to which the English constitution is 'the collection of legal rules and non-legal rules which govern the government in Britain'. Or allow me to quote, as another instance, Jennings's definition, according to which a constitution is 'the document in which are set out the rules governing the composition, powers and methods of operation of the main institutions of government'.[12] The peculiar feature of these definitions is not only the unspecified use of the term 'rule' (even games have 'rules': are they the same kind of rules?), but the silence which covers the *telos* of constitutionalism. Actually they are purely 'formal' definitions, in the sense that they can be filled with any content whatever. We are thus faced, nowadays, with this puzzling situation: that the very inventors of the constitutional solution provide us with a definition which amounts to saying that any instrument of government, any 'traffic rule', is a constitution. And this according to the most authoritative yardstick. It is not astonishing, therefore, if the Continental theory of constitutionalism has shown, since World War II, very little evidence of recovery. For whenever somebody claims that 'constitutional dictatorships' have proved that *garantisme* remains the core of constitutionalism, one is likely to be confronted with this reply: why should we be concerned with this problem more than the English?

My reply is – and this is the second point that I wish to stress, having reference to the English habit of understatement – that one must be very careful about importing the British constitutional textbooks. They have not been written for export. They have been written for a happy people whose constitutional system is liable to work nicely anyhow. If the alien scholar does not make explicit what often remains implicit, and does not say what is often left unsaid, the British case easily lends itself to bearing false witness.[13] Of course, English liberties remain protected even if English constitutionalists forget to mention that this is why Britain has a constitution. But the case might be very different elsewhere.

III

The foregoing *mise au point* has a bearing on what will follow, but it certainly does not alter the fact that nowadays 'constitution' has become an ambiguous term, covering two very different meanings: a strict, substantive meaning (the *garantiste* meaning), and a formal, cosmic meaning. It follows from this that whereas in the nineteenth century a question such as, 'What is the role of a constitution in a political system?', could be answered without asking first, 'What is a constitution?', this is no longer the case.

It is equally clear – at least to me – that if a constitution is defined as 'any way of giving form to any State whatever', then the question 'What is the role of a constitution in a political system?' either cannot be answered, or can be answered only country by country, and even then in a very uninteresting and banal way. For in this case the answer is that the constitution plays no role, properly speaking: It is only a shorthand report which may describe – assuming that the constitution in question is applied – the formalisation of the power structure of the given country. (This is not to deny that the conclusion, 'The constitution plays no role', may not be reached also after applying the standards of the *garantiste* type of constitution. I am simply pointing out that the query is relevant only if we consider the constitutions which are supposed to play a role.)

On both accounts, then, we have to make up our minds and to look for a convincing answer to the question, 'What is a constitution?'

In the first place, we may say: 'This is what I understand when I use this term.' This kind of answer, however, leaves untouched a very basic point, namely: What is the truth-value of this definition? Is it only a 'stipulative definition'?

Therefore, and in the second place, we may also find it necessary to say: 'The term constitution *ought* to be used in this way.'

In the third place, we may have recourse to a classification. We may say: given the fact that we are now using the term constitution in very different meanings and to cover altogether different referents, then let us speak specifically of this and that type of constitution.

These three lines of approach are not mutually exclusive. The contrary is true. But in most cases political scientists either follow solution number one, or solution number three, or both. At any rate, they usually by-pass step number two, that is, they usually refrain from saying that the term constitution *ought* to mean this and not that. I wonder, however, whether the 'ought' can really be avoided.

When general agreement prevails – as was the case with 'constitution' in the nineteenth century – 'ought-propositions' (if I may call them so) are superfluous. In this case one need not discuss whether a definition of 'constitution' is merely a private convention of the stipulator (i.e. a stipulative definition), or a lexical kind of definition,[14] or something more, i.e a definition having warranted truth-value. Agreement automatically settles all these problems. But we live in a world – let me stress – of terminological cleavage in which the 'war of words' plays a very important role. Our first reaction has therefore been to make some fine distinctions (e.g. the one between stipulative and lexicographic definitions), and to have recourse to a classification of the various meanings and referents of the word constitution.

On second thought, however, one is bound to discover that the stipulative approach is only a fire escape (and an unsafe one at that) which actually helps the fire to grow, and that classifications cannot, *per se*, solve our problems either. Classifications are only *prima facie* a neutral device. Classifications have a purpose in view, and differ according to their purpose. Therefore, classifications cannot replace definitions. They follow a definition of constitution, and imply that we have previously taken a stand *vis-à-vis* its nuclear meaning, so to speak. And this stand cannot be provided by the stipulative approach.

According to this approach, the speaker says: I propose to use the term constitution in this sense, and this is *my* definition of it. The emphasis is laid on the 'my', for stipulations are an arbitrary choice. As Richard Robinson asserts, 'In stipulation we freely make any word mean anything we choose.' This is hardly a way out, however. Let us assume, for instance, that one meets with the following definition: a constitution is the will of the sovereign. In some cases one could hardly object that this definition is factually false. According both to the *Führerprinzip* and to actual practice, it would be quite realistic to say that during Hitler's decade the German constitution amounted to the will of the *Führer*. And Stalin's constitution can well be resumed in similar terms. As these extreme but very relevant examples go to show, if one wishes to retort that Hitler's or Stalin's regime was not a constitutional regime, this cannot be done on the ground that the sentence, 'The Soviet constitution was the arbitrary will of Stalin', is factually false, but on the ground that one does not accept the meaning 'will of the sovereign' as a permissible meaning of constitution: i.e. because one refuses the definition.

Yet, if definitions are just arbitrary and private stipulations of each speaker, on what grounds may one discuss the definitions of constitution which differ from his own stipulation? Clearly, one can write a book of 500 pages to show that Hitler's or Stalin's so-called constitution amounted to a non-constitutional system; anybody can simply dismiss it by saying: This is your stipulation, but I have mine, and that is that. How easy, and how convenient!

I cannot develop here the criticism of the conventional approach.[15] In my view definitions are not private conventions of each speaker, but (whenever they have a historical referent) storehouses of past experience shaped by former practice. We bring them up to date, but their ultimate truth-value lies in the fact that they tell us how to behave *as experienced people* in matters regarding which each generation starts by having no experience. And this is precisely the reason why we are not condemned to a solipsistic conventionalism, and why we are entitled to discuss and to look for a 'true definition' (touching on a basic meaning, of course) of

equality, of liberty, of law, of constitution, etc. In particular, *the* definition of constitution which has objective worth is the one that appears to be the outcome of a long and painstaking process of trial-and-error concerned with the question: How can we be governed without being oppressed?

Implicitly I have already indicated why I attach a great importance to whether the sentence 'This is the meaning of the term constitution' has a normative undertone or not. For if the speaker takes the view that his definition cannot be phrased as a normative proposition – i.e. in the form, 'This is also what "constitution" ought to mean' – then, no matter how unwittingly, he is surrendering to the stipulative and conventionalistic approach.

To be sure, one may not be concerned with long-range consequences of this kind. Nevertheless, we must decide here and now the test, or the criteria, according to which we declare: This is the meaning given in this text to the word constitution. Now, are we going to ascertain what 'constitution' means in the twentieth century by asking the Russians, the Chinese, the Egyptians and so forth, or by inquiring what the Italians and the Germans were taught to believe a constitutional system to be under Mussolini and Hitler? If so, then I grant that there is no 'ought' about it. But otherwise (that is, if we do not believe that knowledge follows a majority principle), then the 'ought' is inevitably there – no matter whether openly or not – for surely we are not speaking of what 'constitution' actually means to all the people of the world in the twentieth century, but of a correct meaning which we find advisable to propose for future observance.

I am acquainted with the sophisticated methodological reasons that are being brought forward to explain why the form 'ought' should be expelled from the realm of science, and in particular of political science. Yet, my guess is that future historians will find a much simpler explanation for our 'ought complexes'. They will say that we have been unable to meet the challenge of the environment; and rightly so, I believe. For since the beginning of time the people entrusted with the development of learning have followed this very simple rule: to use the form 'is' when they agreed with general usage, and the form 'ought' when they did not. Yet now we refrain from saying 'This is what the term constitution *ought* to mean', even though we are well aware of the fact that 'constitution' no longer bears a common acceptance and even though we are in a good position to realise that a situation of ambiguity and confusion is being deliberately fostered by political double-talk and insincerity, with the precise purpose of deceiving the audience.

Let there be no misunderstanding. I am not advocating a behavioural 'ought', but a terminological and logical 'ought'.[16] The meaning of 'constitution' is neither an arbitrary stipulation, nor something to be discovered in the 'popular mind' of semi-literate majorities. At least, if we believe that rational discourse and an intelligible vocabulary are to be maintained, then the meaning of 'constitution' has been decided by historical testing, and has to be ascertained by adequate information and sound thinking. And if these criteria bring us to a definition which no longer receives common acceptance, then it is not only right but necessary to say: This is what the term constitution *ought* to mean (for otherwise it is meaningless, or superfluous, or deceiving, or whatever). This does not imply that people are being requested to respond favourably to a certain pattern of constitution defined as the 'true constitution'. I am only saying that terminological clarity is a basic requirement for any science, and that political scientists in particular have to take stock of the fact that the vocabulary of politics tends to be used for the purpose of beguiling the listener. I am not advocating, therefore, the preferability of one type of constitutional *telos* in relation to another. I am simply saying that it is a scientific requirement to discuss whether it is proper to use 'constitution' where, in order that the public (and even, at times, the expert) be not deceived, we should *not* use this term.

IV

The objection will be that even if we reject the stipulative approach, even then 'constitution' is, by its own constitution, an ambiguous term.[17] That is, even if one agrees that constitutions have been shaped by historical experience in order to protect the freedom and the rights of the power addressees, the fact none the less remains that constitution has always been a Janus-faced concept, hovering as it were between the idea of 'political order' on the one hand and of 'limit' on the other. I would, however, disagree with this statement.

The history of the word in its modern meaning only begins in the eighteenth century.[18] Let us not confuse the noun with the

concept, or, we might say, the Latin *constitutio* with our 'constitution', a homonymy with a homology. In its present-day conceptualisation, 'constitution' only emerges, perhaps, with Bolingbroke,[19] and the term really gained ground and acquired a definite connotation only in America during the years 1776–87. (The French did not receive it directly from England, but from the Philadelphia Convention. This is not surprising for it was Paine, not Burke and the English writers in general, who gave the first explicit, complete account of the modern concept.[20])

Now, it is undeniable that the whole of the American tradition has understood 'constitution' as a means for 'limited government'. The same is true, from the outset, for French constitutionalism (should one be reminded of article 16 of the Declaration of Rights of 1789? Or, for the following period, of Benjamin Constant?). And I have previously explained at length why I firmly hold – despite the possibility of exhibiting numberless quotations to the contrary effect – that English constitutionalism belongs entirely to the same tradition: the British advocate, and have advocated from the time of Glanville and Bracton, an 'unlimited government' just as much, or, rather, just as little as the Americans.[21]

It is not unsafe to conclude, therefore, that, with the decline of the age of Absolutism, people began to cast about for a word which would denote the techniques to be used for controlling the exercise of State power. This term turned out to be (Americans decided this issue) 'constitution'.[22] And 'constitution' was in no way born as a Janus-faced concept. The term was re-conceived, adopted and cherished not because it merely meant 'political order', but because it meant much more, because it meant 'political freedom'. We may put it thus: because it denoted *the* distinctive political order which would protect their liberties: or – to paraphrase Friedrich's felicitous wording – because it not only 'gave form' but also because it 'limited' governmental action.[23]

The idea that this specific, *garantiste*, meaning derives from a pre-existing wide, unspecified meaning, is an optical illusion, which has been suggested – I believe – by the commentators and translators of Aristotle. As is well known, Aristotle's term was *politeía*, and, surely, *politeía* is difficult to translate. So occasionally, the authors having Plato's and especially Aristotle's writings in mind, found it expedient to render it by the term constitution. This, however, happened only occasionally (the relevant example is Montesquieu)[24] until the time when 'constitution' acquired a specific meaning. It was only at this stage that 'constitution' came to be used consistently as the proper equivalent of *politeía*. Wrongly, to be sure. For *politeía* only conveys the idea of the way in which a polity is patterned. And if Aristotle meant by *politeía* the ethico-political system as a whole, we cannot infer from this that 'constitution' has in Aristotle a loose meaning. The only correct conclusion is that Aristotle has been mistranslated. For to us 'constitution' means – a frame of political society, *organised through and by the law*, for the purpose of restraining arbitrary power. And surely nothing resembling this concept was in the mind of Aristotle.[25]

V

Having established the frame of reference (the historically valid 'nuclear meaning' of constitution), we may now turn to the problem of how we should go about classifying the variety of present-day so-called constitutions. I shall deal swiftly with this problem, for I agree very much (in substance, even though not in terminology) with the threefold classification suggested by Loewenstein.[26]

Basically we are confronted with three possibilities: (i) *garantiste* constitution (constitution, proper); (ii) nominal constitution; (iii) façade constitution (or fake constitution).

I call 'nominal' the constitutions that Loewenstein labels 'semantic' (a difficult, and perhaps not quite appropriate labelling). Apart from this semantic divergency, I entirely underwrite his description of the specimen: 'The constitution is fully applied and activated, but its ontological reality is nothing but the formalisation of the existing location of political power for the exclusive benefit of the actual power holders.' (Nominal constitutions are therefore 'nominal' in the very simple sense that they bear the 'name' constitution. This amounts to saying that nominal constitutions are merely organisational constitutions, i.e. the collection of rules which organise but do not restrain the exercise of political power in a given polity.[27] Actually, nominal constitutions do not really pretend to be 'real constitutions'. They frankly describe a system of limitless, unchecked power. They are not a dead letter. It is only that this letter is irrelevant to the *telos* of constitutionalism.

The façade constitutions are different from the nominal ones in that they take the appearance of 'true constitutions'. What makes them untrue is that they are disregarded (at least in their essential *garantiste* features). Actually they are 'trap-constitutions'. As far as the techniques of liberty and the rights of the power addressees are concerned, they are a dead letter.

I suppose that the labelling 'façade constitution' will be criticised on the ground that it implicitly rules out the possibility of an educational purpose. But I do not think that a fake constitution has an educational *purpose*. It *may* turn out that it has an educational *effect*. But this is a very different matter. We are not historians dealing with past events and looking for their *a posteriori* justification. We are concerned with actual will and actual doings. And to credit with an 'educational role' the instances that makes forecasts all the more dubious (as in the case when we are confronted with discretional power) is very much like toying with wishful thinking. Moreover, and in point of principle, education is not the purpose of constitutions. I mean, this is not a sufficient (or sufficiently pertinent) criterion for singling out a special class of 'educational constitutions'.[28] A constitution may contain, to be sure, statements of 'aspiration', and thereby provide 'incentives', even pedagogic incentives. However, this occurrence does not indicate an essential constitutional feature; it merely indicates a possible content of any type of constitution.

There is often a considerable overlapping between nominal and façade constitutions. The distinction is nevertheless basic, for the two cases are indeed very different. Nominal constitutions actually describe the working of the political system (they do not abide by the *telos* of constitutionalism, but they are sincere reports), while the façade constitutions give us no reliable information about the real governmental process. In most cases one can clearly perceive, despite the overlaps, which is the prevalent aspect: I mean, whether a constitution is basically nominal or basically a disguise. At any rate, or otherwise, the distinction is serviceable for analytical purposes, that is, for dissecting the component parts of a 'mixed type' (partly nominal and partly fake) of pseudo-constitution.

The distinction is serviceable from another point of view as well, for it goes to show that if we refuse the *garantiste* qualification, then no line can be drawn between 'constitution' and 'constitutional government'. To be sure, if a constitution is not applied, this is an instance of constitution without constitutional government. But this argument applies only to the façade constitutions. What about the other possibility? In the case of nominal constitutions we no longer have a 'prerogative state' that replaces *de facto* the 'legal state'; we have instead a prerogative state legitimised by the constitution. In this latter case, then, we are confronted with a government which, by definition, will always govern according to the constitution. On what grounds may we maintain that this is not a 'constitutional government'? Certainly not on the grounds of logical syntax. Therefore, either we adhere to *garantisme*, or the distinction between constitution and constitutional government really becomes meaningless.

Some troublesome problems arise when one focuses further attention on: (i) the *décalage* between the written and the living constitution, and on (ii) the frequent disregard of some of the constitutional provisions. In this connection the point can be made that a clear-cut distinction between real constitutions and façade constitutions is hardly realistic, since the real ones too come to differ widely, in practice, from their original formalisation, or may not be fully activated (the former being usually the case with old constitutions, and the latter with recent ones).

Personally I am not dismayed by the first indictment. If a constitution is written, then, with the passing of time, the formal document and the living constitution inevitably come to be related much as the past is related to the present. (In this sense, the written constitutions too become, in part, non-written.) However, as long as the spirit and the *telos* of the original document are maintained in the new circumstances, the *décalage* only affects the myth of a 'fixed constitution'; and the American experience goes to show, if anything, that written constitutions can endure despite the anti-historical assumption upon which they have been conceived.

On the other hand, the remedy for this could hardly be found in a non-written constitution. The drawbacks and the dangers of this solution are so serious that only the British can afford the luxury of not formalising their constitution. And it remains questionable whether it is really true that the British constitution is unwritten. (I would be tempted to say that it is 'written differently'.)

The thorny point is instead the non-fulfilment of constitutional provisions not because of the time factor, because they have gradually become outdated, but with reference to norms that have never been activated owing to the unwillingness of the executive or of the legislative body to give them life. This problem cannot be dismissed lightly, if we consider that 'delinquencies in the application' of the constitution (as Loewenstein calls them) are rather frequent in most countries. It is safe to ask: 'Why?' Is it because the constitutional spirit (as well as the corresponding type of the constitutional gentleman) is withering away? Or is it because of other reasons?

It is well to remind ourselves that most countries have a recent constitution, either because they have rewritten their previous charters, or because they have started anew. And contemporary constitutions are, as a rule, bad constitutions – technically speaking.[29] They have come to include unrealistic promises and glamorous professions of faith on the one hand, and numberless frivolous details on the other. Some of them are by now so 'democratic' that either they are no longer constitutions (for a constitution limits the 'will of the people' concept of democracy just as much as it limits the will of the power holders),[30] or they make the working of the machinery of government too cumbersome for government to work, or both. Under these conditions, non-application may well be a remedy for inapplicability. I mean by this that it may also serve the purpose of saving the constitutional system as a working system.

So we should not react dramatically in all cases, but only in those in which the disregard actually has a direct bearing on the withering away of the constitutional *telos*. This implies that we can no longer partake, in all instances, of the strict juristic view that '*all* the Constitution *must* be applied at whatever cost'. Personally speaking at least, I take the view that we should always ascertain whether the non-application affects the machinery of government in its *garantiste* aspect and the basic purposes of constitutionalism, or not. In the former case it is proper to speak of 'delinquencies'; in the latter I confess that I remain quite indifferent. And if we follow this rule of thumb, the difference between 'real constitutions' and 'façade constitutions' remains very firm, despite the fact that even in the context of the real constitutions a number of provisions may not be activated.

VI

Summing up my line of argument, I have tentatively submitted the following theses:

(1) That the word constitution has been re-conceptualised, if I may say so, for a specific purpose, and more precisely to denote a distinctive and unprecedented 'technique of liberty' (as Mirkine-Guetzévitch would have said), the features of which had already materialised clearly enough when, for instance, Montesquieu was visiting England in 1730.

(2) That this purpose has often been obscured by the complexity of the original model (the English constitution) on the one hand, and by a somewhat polemic isolationism of the British scholars on the other hand. Despite these unfortunate circumstances, however, the fact remains that the idea of limit is basic to the English prototype just as much as to the American and the French subsequent models.

(3) That the view that the *garantiste* meaning of 'constitution' was preceded by a loose, formal meaning lacks historical proof, for this view actually goes back to a mistranslation, or to a casual way of referring to Aristotle's term *politeía*.

(4) That the equivalence 'constitution = any state form' is therefore not the older but indeed a recent loosening-up of the concept, reflecting either the juridic illusion of attaining a 'purified', universal depoliticised right, or reflecting the purpose of exploiting the word constitution as a trap word.

(5) That on both accounts it should be pointed out that either the term is used in its specific *garantiste* meaning, or it is a meaningless (and deceiving) duplicate of terms such as organisation, structure, form, pattern, political system and the like.

(6) That, according to this view, the variety of so-called 'constitutions' of our time may be classified in three categories: real, nominal and façade.

(7) That the existence of nominal constitutions implies that either we abide by *garantisme*, or we can no longer draw a line between 'constitution' and 'constitutional government'.

(8) That the political scientist, in particular, is responsible for ruling out insincere language, for two reasons: the general

reason being that the language of the observer is required to improve (whenever necessary) the language of the observed; while the specific reason is that terminological house-cleaning is all the more necessary whenever the political scientist is confronted with terminological distortions that are deliberately fostered with a view to deceiving the audience.

Speaking of political scientists, a final question should perhaps be raised, namely: why should we, *qua* political scientists, be concerned with the constitutional problem?

Actually, constitutionalism is not a traditional subject-matter of political science. To the extent that political scientists have been concerned with it, they have either leaned on a previous legal training of the individual scholar, or they have shown a tendency to underplay the role and importance of constitutional checks as compared with social and political pluralistic checks.[31] In the first case it could be said that their treatment of the topic constitutionalism belongs to the 'old' political science. In the second case it can be shown, I believe, that the 'new' political science is hardly equipped to discuss the problem at all. For one thing, political science is keen about informal processes, not about formal structures. In particular, the new generation of political scientists distrusts concepts having a high degree of abstraction (such as State), concentrates very much on elementary grass-roots units (such as groups), and has developed a keen feeling for dynamics, for change, for 'something doing', to use Bentley's telling expression. Now, clearly, from this platform constitutions appear to be: (i) the juridical counterpart and support of the abstract entity, 'the State'; (ii) certainly not the 'raw material' of politics; and (iii) the expression *par excellence* of a typically static view of the political process. (The foregoing are, of course, very broad and very rough generalisations.)

If these underlying premises are duly weighed, the present-day mood of scepticism towards the efficacy of constitutional devices and mechanisms was only to be expected. However, where is the error? In the purview of the observer, or in the constitutional solution? All in all, my feeling is that contemporary political science has not found, as yet, the strategic juncture in which and from which it can make a positive contribution to the understanding of constitutionalism. My own tentative suggestion is that, perhaps, this strategic point is to be found in the 'role theory' approach,[32] that is, by examining the constitutional solution as a pre-ordained, binding technique of *role-enforcement* (in its difference from a spontaneous role-taking.)

Within the limits of a preliminary perusal, I cannot push the argument farther. I leave it as an open suggestion for discussion that the query 'What is the role of a constitution in the political system?' could perhaps be fruitfully approached from the following angle: what is the impact, or the role, of a constitution *vis-à-vis* the role-taking of the power holders? That is, does it help to enforce, and if so to what extent, a desired 'role performance' upon the persons in office?

Whatever the merit of this suggestion, I do feel that political scientists should go back with a fresh look to the vital issues that constitutional lawyers have been neglecting. For the experience of the last thirty to forty years has shown, particularly in Italy and in Germany, that when a political problem – and constitutionalism is, inescapably, a juridical solution of a political problem – is depoliticised, the real consequences of taking the juridical 'neutral' attitude are (no matter how unwittingly) political. I mean that when the time of trial comes, one discovers that what the 'pure' jurists have really been doing – under the shield of their juridical indifference to meta-juridical matters – was to pave the way for allowing unscrupulous politicians to make a discretionary use of power under the camouflage of a good word. Politics cannot be taken out of politics, so to speak.

Notes

1. Despite Jellinek's contrary opinion. According to Jellinek, 'constitution' derives from the Latin phrase *rem publicam constituere*. However, no historical continuity appears to exist between *constituere* and 'constitution' (see below, notes 14, 18), and this derivation is bound to give us a misleading start.

2. K. C. Wheare, *Modern Constitutions* (London, 1960).

3. W. I: Jennings, *The Law and the Constitution* (London, 5th edn., 1959).

4. Bagehot is no exception when he asserts that 'a new House of Commons can *despotically* . . . resolve' [my italics]. (*The English Constitution*).

5. This is to deny that there is a substantial difference between the British and the Continental principle of parliamentary sovereignty. If the term of reference is the American Congress, then, of course, there is a good deal of difference.

6. A comparison between Jennings's classic text and Vishinsky's *The Law of the Soviet State* (New York, 1951) would be to the point. According to Vishinsky, Stalin's Constitution draws 'ever broader masses of the people into the government of the State', constantly strengthens 'the bonds between the apparatus of authority and the people'; and Soviet constitutions in general 'confirm genuinely democratic rights and freedoms', 'establish and emphasise material guarantees', etc. (pp. 88–9).

7. In 1860, the formula for the Italian plebiscites said only: 'Do you want to enter the constitutional monarchy of King Victor Emanuel II?' Evidently, the mere adjective 'constitutional' was assumed to be understood as implying all the difference.

8. Of course the query, 'What do the guarantees include?' (e.g. a certain technique of allocation of power, a bill of rights, the rule of law, judicial review, etc.) receives different, complex and changing answers. This is all the more reason for adopting a general term, reminding us both of the goal, and particularly of the fact that unless we think that *somebody needs protection against somebody else*, there is no point in being concerned with constitutionalism.

9. The extent to which political terminology is subject to this kind of abuse can be exemplified by the very term 'politics'. In the Middle Ages the expression *dominium politicum* meant (in contrast to *dominium regale*) government deriving from, or given consent by, the people (i.e. by the *politēs*, the inhabitant of the *polis*). That is to say that *politicum*, or *police* (in French), was coined – having reference to the Greek root – as a good word. Nowadays this originally pleasant word denotes the most unpleasant reality of politics: those who are entitled to arrest us. Let us hope that 'constitution' may not have a similar destiny.

10. Pellegrino Rossi (1787–1848) was professor of law in Bologna in 1814, the framer of the 1832 project of reform of the Swiss constitution, and subsequently a celebrated professor of constitutional law at the Collège de France in Paris.

11. The wording varies, but this was, insubstance, the invariable gist of their argument.

12. Or, where no such document exists, simply 'the rules determining the creation and operation of governmental institutions' (*op. cit.*).

13. Or else to bear no witness. Thus in F. A. Hayek's *The Constitution of Liberty* (Routledge, 1960), the 12th chapter bears the title 'The American Contribution: Constitutionalism'. To be sure, Americans have made an outstanding contribution to constitutionalism. But to say more than this is saying too much, even though Hayek's conclusion is hardly surprising if one is not cautioned against the British mood of understatement.

14. See R. Robinson, *Definition* (Oxford, 1954); M. Cranston, *Freedom: a new Analysis* (London, 1953).

15. G. Sartori, *Democratic Theory* (Detroit, 1962).

16. The form 'ought' is not necessarily related to ethical imperatives.

17. I am no longer referring to the strictly juridicial approach mentioned previously, but to the overall issue of constitutionalism as discussed in the history of political thought and in philosophy of law.

18. Cicero's use, in *De Republica*, I 45, 69, was quite casual, and left no trace in the following literature. Actually, this much-quoted Roman precedent cannot claim precedent status. This is so true that the fifteenth- and sixteenth-century commentators rendered Cicero's *constitutio* with the terms *status publicus* (Jean de Terre Rouge), *status Reipublicae* (Bodin) or *politicum* (Fortescue), *police* (Seyssell).

19. I say 'perhaps' on account of McIlwain's reservation: 'Bolingbroke in fact is only restating views as old as the *Politicus* of Plato. . . .'

20. In Burke 'constitution' was used much more loosely as a synonym for 'commonwealth' (*Thoughts on the Cause of Present Discontents*), for 'engagement and pact of society' (*Reflections*), for 'frame' and 'pattern' (*Speech on Reform*); all in all, it meant the substantive principles deriving from the British political tradition and to be inferred from the working of the institutions.

21. Cf., e.g., Burke: 'In [our] Constitution . . . I feel both that I am free, and that I am not free dangerously to myself and to others. I know that no power on earth, acting as I ought to do, can touch my life, my liberty, my own property' (*Speech on Reform of Representation*). Let it be noted that, according to the letter of the constitution, in 1782 the doctrine of the omnipotence of Parliament was already established. Was Burke wrong, then? Or should we not question, instead, the validity – for a constitution largely based on conventions – of the 'literal' approach?

22. In this connection, recall that in the years of the Commonwealth and the Protectorate (1649–60) the English made several attempts to establish a written constitution. However, they never called these documents 'constitution': they made recourse to terms such as covenant, instrument, agreement, model, paramount or fundamental law.

23. C. J. Friedrisch, *The Philosophy of Law in Historical Perspective* (Chicago, 1958).

24. In the famous chap. 6, bk XI, of the *Esprit des Lois* the term *constitution* appears only in the title.

25. I grant that there are a few passages in Aristotle's *Politics* in which the term *politeía* specifically refers to the way in which the *polis* magistracies are ordained. Cf., e.g., 1278 b. However, this meaning appears in passing, and it cannot be denied, I believe, that (i) the ethico-political all-embracing meaning of *politeía* is by far the prevalent one, and, on the other hand, (ii) the term never appears in Aristotle in our juridical meaning, that is, in the sense of being an institutionalised way of controlling the dynamics of political freedom. Actually, I find it quite absurd to speak of a Greek 'constitutionalism'. Only the Romans indirectly tackled the problem, and only from the 'rule of law' angle (as Jhering and Bryce have remarked), that is, very much in the same way in which – according to Dicey – the principles of the English constitution are inductions and generalisations resulting from judicial decisions determining the rights of private persons.

26. K. Loewenstein, *Political Power and the Governmental Process* (Chicago, 1957).

27. It is often held that the mere ordering, the mere existence of a definite and stable 'form', is in itself a 'limit'. This may well be – but

not a *garantiste* kind of limit. Armies are usually well ordered, but this fact does not necessarily protect subordinates *vis-à-vis* their superiors.

28. Car accidents may serve the purpose of improving our driving. Would this be a reason for creating the category 'educational accidents'?

29. Contemporary constitutions are being said to improve on the former ones in that they are no longer 'negative' but 'positive'. If positive means that they are also an instrument for social and economic policy, then let us be happy with this positive development. Under one condition, however: that the follow-up, i.e. the 'economic' tail, should not eat the 'political' head.

30. Among other reasons, because a constitution cannot effectively limit the will of the power holders if they can outflank constitutional impediments by making direct appeals to the will of the people.

31. This is not to deny that pluralism is a requisite condition for the proper working of constitutional systems. My own view, however, is that the relationship between societal pluralism and constitutional patterns of behaviour is a two-way relationship, at least in the sense that it is very important, for pluralism, to have the constitution on its side.

32. I have in mind, e.g. the approach devised by Heinz Eulau and associates in a related subject, see Wahlke, Eulau, Buchanan, Ferguson, *The Legislative System: Explorations in Legislative Behavior* (New York, 1962).

Much as can be learned from formal structures, there is far more to government and politics than constitutions and such institutions as legislatures, executives and courts that are described and delimited in constitutions. Informal institutions – parties, interest groups, factions or other bodies – operate sometimes within the limits of the formal structure and sometimes in defiance of it. The form and working of these informal institutions can be, and frequently is, affected by the formal structure, but conversely they determine to a great extent how the machinery of government operates and to what end its power is devoted. The two articles in this chapter present the results of a comparative analysis first of parties and secondly of interest groups. The first appeared originally as a concluding survey in a book that included studies of the parties in a series of countries in Western Europe, the Commonwealth, the United States, Japan and the Soviet Union. The second reports the deliberations and conclusions of a body of scholars concerned with the comparative study of interest groups.

TOWARDS A COMPARATIVE STUDY OF POLITICAL PARTIES

Sigmund Neumann

I. A PRELIMINARY DEFINITION

A DEFINITION of 'party' might as well begin with its simple word derivation. To become a 'party' to something always means identification with one group and differentiation from another. Every party in its very essence signifies *partnership* in a particular organisation and *separation* from others by a specific programme.

Such an initial description, to be sure, indicates that the very definition of party presupposes a democratic climate and hence makes it a misnomer in every dictatorship. A one-party system (*le parti unique*) is a contradiction in itself. Only the coexistence of at least one other competitive group makes a political party real. Still the fact remains that the term has been widely used by modern autocrats, and for a very obvious reason: to keep the semblance of a 'people's' rule' in their post-democratic dictatorships. But it is also true that even the totalitarian party depends upon a functioning opposition. If one does not exist, it must still be assumed by the dictators, since under monolithic rule the dictatorial parties must constantly justify their existence in view of the ever-present threat of a counter-revolution, hidden or imaginary though its organisation may be. The opposition party is the *raison d'être* of the dictatorial movement and its all-pervasive controls through institutions, propaganda and terror.

Thus parties must prevail under the total structure, too. Yet here again, as in all other strata of the political pyramid, divergent political systems are hidden behind the same nomenclature. Just as 'leaders' and 'lieutenants' carry different meanings in democracies and dictatorships, so do 'parties' also.

Between the two extremes of 'democracy' and 'dictatorship' there are wide variations. The British, American, French, Indian and Scandinavian democratic structures differ as greatly among themselves as do the totalitarian regimes of the Soviet Union, Fascist Italy, Nazi Germany and Peronist Argentina. In short, the peculiar character of each party system must be defined in terms of the political order of which it is an integral part, if not its kingpin.

What is common to all parties, beyond partnership in a particular organisation and separation from others, is their *participation* in the decision-making process, or at least the attempt at and a chance for such a mobilisation for action. This ever-present readiness alone makes them political in a genuine sense, for only in their fight for control and in their conscious influence on political forces do parties gain meaning and importance. It is therefore not accidental that the beginning of modern political parties is closely tied up with the rise of a parliament. When political representation broadens and a national forum of discussion develops, providing a constant opportunity for political participation – wherever those conditions are fulfilled, political parties arise. This happened in England in the revolutionary seventeenth century; in France on the eve of the great Revolution of 1789; in Germany around 1848. Even where contingent influences may create political groups of an awakened intelligentsia, as in nine-

teenth-century Tsarist Russia, they assume political dimensions only where some degree of participation is made possible.

While the hour of birth for the political party in every nation can well be defined by such a simple derivation from its original meaning, the same is true of the critical period of a party system. By its very definition, 'party' connotes not only the coexistence of different competing entities, with their characteristic partnership, separation and participation, but also a fourth feature, most significant and yet often forgotten – the essential inclusion of every separate group as *a part in a whole*. Only where the specific interests of parties are embedded in a common whole does the political struggle not lead to disintegration of the entire group. Only when essentials uniting the political adversaries are constantly reaffirmed can differences be balanced. Just as children are ready to accept the rules of the gang only as long as they are willing to continue their group life, so can sacrifices be asked from each political opponent as long as the preservation of the community seems worth while. A common field of activity, a basic homogeneity, a common language, are presuppositions for a functioning party system. Such a common basis alone makes compromise, sacrifice and even defeat bearable. Wherever this body politic becomes questionable, the crisis of parties seems the necessary result. The viability of a party system becomes a test for the stability of a social and political order. The strength of the Anglo-American party system is founded largely upon a basic national unity which makes the differentiations of political groups 'differences in degree but not in kind'.

This interdependence of the fate of political parties with the fate of the national whole is the very result of their 'political' character; for political they are not only because of their claim to political power, but even more so on account of their fulfilment of their political function of integration.

To summarise: we may define 'political party' generally as *the articulate organisation of society's active political agents, those who are concerned with the control of governmental power and who compete for popular support with another group or groups holding divergent views*. As such, it is *the great intermediary which links social forces and ideologies to official governmental institutions and relates them to political action within the larger political community*.

A party's concrete character can be spelled out only in time and space. What 'this buckle' (to paraphrase Bagehot's description of the British cabinet) actually links up depends on a nation's specific constitution (i.e. the system of its correlated institutions) and its peculiar parallelogram of social forces. It is only in such a substantial situation that the functions, structure and strategy of parties can be fully revealed.

II. FUNCTIONS OF POLITICAL PARTIES: DEMOCRATIC AND DICTATORIAL

It has often been stated that the primary task of political parties is to organise the chaotic public will. 'They bring order out of the chaos of a multitude of voters' (Lord Bryce). They are brokers of ideas, constantly clarifying, systematising and expounding the party's doctrine. They are representatives of social interest groups, bridging the distance between the individual and the great community. They maximise the voter's education in the competitive scheme of at least a two-party system and sharpen his free choice. The coexistence of a competitor, therefore, is paramount to an effective democratic party system which presupposes that the final compromise will reflect the reasonable decision of a free electorate.

In fact, the basic assumption of democracy is the inevitability of differing views and the free operation of conflicting opinions. 'The true democrat has a suspicion that he may not always be right', as W. Ivor Jennings remarked. Thus the opposition becomes the most important part of the parliament; its members are 'critics by profession'. It is not enough that Her Majesty's Opposition be highly respected by the ruling majority, but often its fruitful ideas are accepted – and, indeed, this is a wise course for the party in power to follow if it wants to remain there. Its political alternative represents not only the looming 'Shadow Cabinet' but also an active participant in actual control.

The open forum of parliament becomes the clearing-house for the policies that a state should follow. The political parties are the proper engine of such continuous plebiscite. They make the voters choose at least the lesser of two evils, thus forcing political differentiations into a few major channels. Yet important as such machinery of political concentration may be, the political services of the parties do not stop at that. What is even more essential, parties

transform the private citizen himself. They make him a *zoon politikon*; they integrate him into the group. Every party has to present to the individual voter and to his powerful special-interest groups a picture of the community as an entity. It must constantly remind the citizen of this collective whole, adjust his wants to the needs of the community, and if necessary, even ask sacrifices from him in the name of the community. Nor can even the so-called 'class parties', which call upon only a specific part of the population, renounce this essential function. The outstanding example of such a class programme, the *Communist Manifesto*, justifies its position with the claim that the united proletariat will represent the overwhelming majority and that its dictatorship will lead to a dissolution of all classes and therewith to the liberation of society as a whole. This second function differentiates the political party from a pressure group. Its specific interests must be fitted into the framework of the national collective. Wherever the policy-making parties do not succeed in this primary task, the modern state is in danger of deteriorating into a neo-feudalism of powerful interest groups.

If the party in a democracy fulfils these two first functions of organising the chaotic public will and educating the private citizen to political responsibility, then it can also lay claim to a third duty: *to represent the connecting link between government and public opinion*. Since democracies are pyramids built from below, the connection between leaders and followers becomes a necessity in the two-way traffic of democracy. It is the major function of the party to keep these lines of communication open and clear. Such a task makes the parties, if not the rulers, at least the controlling agencies of government in a representative democracy.

This crucial position is even more emphasised by the fourth function of a democratic party: *the selection of leaders*. Here as everywhere in a democracy, it is the competitive scheme, the choice between at least two oligarchies, which guarantees the quality of its leadership. Of course, such a selection presupposes an enlightened public, one qualified to make the right choice, and an intellectual climate appropriate to the functioning of democratic parties. Wherever these preconditions no longer prevail, the crisis of democracy is in the offing.

The crisis elements of democratic parties can be well perceived in the rise of dictatorial movements. In fact, dictatorial organisations often grow up within the democratic party system itself. They constitute a state within the state, alienated from its basic principles. And yet their rise is the expression of a basic lack within the society. They can recruit followers because there are those who no longer regard themselves as a part of the predominant society that does not seem to answer their essential desires and needs. No doubt every functioning group has and can bear some 'outsiders' in its midst. So long as they do not represent a considerable number, they do not constitute a serious threat to the existing order. If, however, they succeed in recruiting an appreciable following, then the democratic process enters a critical stage. The rise of dictatorial parties within modern states is a storm signal for the democratic party system. From now on, the argument between parties concerns fundamentals and a fight over ultimate issues. For these integrated political groups compromise becomes increasingly difficult, and so does any coalition with another party.

The main purpose of the fully developed totalitarian parties becomes the fight for a new political order, for a new society. Speaking a different language and living according to a different set of values, the partisans have to be segregated from the political and social body of the ruling class and society, so the party leaders decree. Otherwise the partisans may be enchanted and taken in by the old order, the destruction of which is the essential purpose of the 'guarantors' of the new morrow.

With the rise of the dictatorial party, its competitors in politics necessarily become more inflexible too. Struggle assumes the quality of a religious war; the only possible outcome for any contestant seems to be overwhelming victory or ultimate annihilation. Such a situation explains the revolutionary functions of a dictatorial party before its seizure of power: it is, above all, the *revolutionary vanguard* of the future state.

The functions of the dictatorial parties in power, outwardly at least, do not appear to be different from the four features of their democratic counterparts. They, too, have to organise the chaotic public will and integrate the individual into the group; they equally have to represent the connecting link between government and public opinion and, above all, guarantee the selection of leaders. Yet as their concepts of leaders and followers differ diametrically from democratic ideas, the meaning of these functions changes fundamentally. Organisation of the chaotic will is fulfilled by a 'monolithic control'; integration of the individual

means 'enforcement of conformity'; and, though these tasks are often directed by a 'Ministry of Education and Enlightenment', the maintenance of communication between state and society is assured by a mere one-way propaganda stream from above. True, dictatorships also have to concern themselves with 'public opinion'. They must listen to the 'voice of the people', especially since this is muted under the tyrant's rule. Thus the party serves through its secret agencies as a necessary listening post. Through such diverse services this leviathan apparatus, which claims at the outset to be the party to end all parties, becomes in fact the key instrument of modern totalitarianism.

All three functions, if successfully administered, secure the fourth and crucial purpose: the selection of leaders. Yet it is especially on this level of the creation, preservation and extension of the ruling elite that the basic differences between the party systems become obvious.

Lenin's fight at the turn of the century for a small centralised revolutionary elite, as opposed to the Mensheviks' idea of a loose democratic mass organisation, laid the foundation for a disciplined caste-like order. He thus anticipated what revolutionary parties experienced a generation later when they had to choose between thoroughly revolutionist cadres and a mass following.

Only in a 'revolutionary situation' – i.e. when complete victory or the prospect of impending success brings a rush of adherents to the revolutionary cause – is it possible for such a radical party to win and to hold the masses. Revolutionary parties reckoning with a long struggle can count on only a small elite of unrelenting fighters who do not care for rewards today and who are ready to make the revolution their life's calling. Masses are in need of visible rewards. If they cannot reap the fruits now – or at least have reasonable expectation of doing so in the near future – they will leave the ranks. This fact explains the extraordinary fluctuation in membership to be observed in radical parties everywhere. The Fascist followers of yesterday are turning up today as communism's most reliable fighters; and the reverse may be true tomorrow. Whoever delivers the spoils has the confidence of the fluctuating masses; but if he cannot do so by tomorrow, another liberator will be sought out.

How different from the rank-and-file reaction are the attitudes of the vanguard and the core members of the established movement, taking their orders from the central party organisation, whether it means robbing a bank, organising a party cell or conceiving a new programme. The revolutionary intellectual in his great flight into a future morrow and the revolutionary 'wire-puller' in his minute work, far apart though they be, have one thing in common: they do not expect rewards in this world. They have given their lives to the party, which in return becomes their own life. Discipline is the password of the movement. Said Lenin in 1900: 'We must train men and women who will devote to the revolution not merely their spare evenings but the whole of their lives.' This is the clarion call to the totalitarian party. It can appeal to the professional revolutionary at all times; yet to the masses it may be meaningful only during a revolutionary situation.

The Continental Communist parties in the inter-war period opened their ranks to the masses partly because of their mistaken misapprehension of what they deemed a revolutionary situation (which had passed by 1920). They did not fare too well. In Soviet Russia, of course, they could offer their followers the spoils of the victor. In fact, many people joined the Bolshevik party for such reasons alone, and the early purges of the Soviet regime were partly to free the party from those 'opportunists' (and more often from those whose services were originally useful but which now could be performed by newly recruited 'reliable' followers).

The radical parties in the non-Communist countries did not have to eliminate 'job-hunters' because there were no jobs to be distributed. As quickly as the prospects of revolutionary changes attracted the driftwood of a political reserve army, so did this army move quickly on to new shores which seemed more promising. The fate of the Communist parties outside Russia in the inter-war period hinged on the dilemma between their revolutionary character and their aspirations for mass following, between their uncompromising attitude as a party which did not want to sell the revolution for a 'mess of pottage' and the necessity of offering substantial results at once because the masses would not be satisfied with a promised millennium. They wanted higher wages, social security, political recognition – visible signs of success today. Thus they left the party that was preparing for a revolution not yet in sight.

A mass party cannot survive without foreseeable success. Its final fate becomes a race against time. The surprising staying power of the Communist parties in Western Europe since 1945 seems to contradict such a statement; yet, as Charles Micaud's analysis

convincingly shows, the left-wing radicals have learned some lessons from their earlier failures and have defined more realistically the function and place of the totalitarian party in what they regard as the preparatory stage of an oncoming revolution. In fact, the continued appeal of communism in France and Italy, despite the Marshall Plan and NATO, serves as a warning to the democratic powers in areas where they have not yet succeeded in reintegrating large segments of their national community. It is the claim of the Communist movements that they can give their followers a feeling of belonging, an active faith and a promising direction for the future when the party will possess the whole community and seize the power of the state.

III. TOWARDS A CLASSIFICATION OF PARTIES

Since political parties are mainly concerned with the control of governmental power, the most obvious differentiation would be to distinguish the *in-group* from the *out-group*. Indeed, such a classification may point at some fundamental traits of political strategy and emphasise the advantage of a two-party system for clear-cut confrontation. It would divide the 'haves' and the 'have-nots' in politics. Thus the insider at the controls would be often identified with the *status quo*, the conservative tendencies, while the challenging outsider would usually be the party of change and reform.

One may even go farther and recognise in these two opposing camps a classification which has been frequently applied by Continental writers as the possessor's *party of patronage* versus the defiant's *party of principles* and thereby indicate the danger points of the two parties: the corruption of the 'ins' by power and the irresponsible dogmatism of the 'outs'.

In a wider comparative approach, the degree of proximity to power may also explain interesting variations in national party systems. Wherever parties are called upon for political decision-making, as in the Anglo-American democracies, they may well emphasise the day-by-day *expediency interests*. On the other hand, in nations where the parties have played only a subordinate role or where they are purposely kept from political key positions by the government, as was the case during the greater part of German history, parties may easily retreat to the fundamental principles of an all-inclusive 'faith movement' (*Weltanschauungs-* or *Glaubens-Parteien*).

These different national positions may also suggest a division of parties of *platform* and of *programme*. One may, however, observe that such a strict cleavage has become less significant in recent decades. The more the German parties have been drawn into responsible politics, especially in the contemporary attempt at a democratic government, the more they are forced to make daily decisions along lines of concrete interests. Contrariwise, the Western democracies have been increasingly confronted with fundamental issues of international reorientation, national planning and individual soul-searching – all of which demand a basic ideological outlook. This interesting *rapprochement* of seemingly contradictory systems will have far-reaching consequences and deserves our most careful consideration. At this point we must mention that the traditional lines of demarcation between the parties of expedient interests and those of fundamental principles have been blurred.

In the light of these developments another classification is losing its sharp contours, the suggestive division of *parties of personages* versus *parties of programmes*. Although this dichotomy has reappeared today, but in a different version, in the contrast between dictatorial and democratic rule, between personal and institutional government, parties of personages are of the past. Max Weber's three stages of party development from aristocratic cliques to party of notables to plebiscitarian democracy point to a general trend in the Western world. The British party growth is a case in point, as Samuel Beer's careful study clearly shows.[1] On the Continent, leaving aside preliminary aristocratic factions, the early period of parliamentary development, with its emphasis on local representation, was characterised by notables (*Honoratioren*) giving their name and decisive mark to political groups (as in the German parliament of 1848, the centre and right-wing factions in the French Third Republic, and in pre-Fascist Italy). Yet today, when nation-wide public issues of class, religion, nationality and international affairs constitute the dividing line of political parties, their leadership is forced to comply with the overall programme. Thus the parties became nationalised even in such traditionally decentralised structures as the United States, as E. E. Schattschneider's chapter[2] pointedly describes (though personalities may still be crucial in making or breaking a party's hold on a nation, as the recent

presidential elections in the United States abundantly testified).

The reality of modern politics represents a much more complex picture than is suggested by the simple array of insiders and outsiders, of parties of patronage and parties of principles, of expediency interests and *Weltanschauung*, of personages and programmes. Such precise but utterly imaginary partitions fail to reveal the inner dynamics and tensions of a functioning democracy. In fact, it is the inexhaustible mixture of all these elements that comprises the life of modern political parties – and perhaps escapes any rigid classification.

A more modest attempt at systematisation simply follows the number of contestants in the political field and speaks of one-, two- or multi-party systems. But some important facts of political organisation and control can be derived from such a classification.

A careful analysis of the causes and consequences of the two-party system versus the multi-party regime must probe into the historical circumstances, social structures and institutional arrangements which underlie the divergent national political settings. While it is generally agreed that a two-party rule promises greater efficiency for the democratic process, as the British and American cases prove, it is equally obvious that such a political setting cannot be easily transplanted into national communities which do not meet its preconditions and which therefore, by no mere oversight or accident, possess a different political organisation.

Historical precedents may suggest the following favourable circumstances for a two-party development: social homogeneity, political continuity, an early sanction of responsible political parties striving for political control and their orientation at one elective office (the United States presidency, the British premiership) as the desired prize.

Wherever fundamental cleavages in social structure evolve and continue to exist because of differences in nationalities, regions, religion or class which are often fostered by outside influences like *irredenta* movements and revolutionary internationals; wherever political revolutions coincide with great social transformations, as in France, Central and Eastern Europe and the Near and Far East; wherever a controlling elite, through the divide-and-rule device, prevents parties from fulfilling their genuinely political functions of presenting clear-cut policy alternatives, as in Bismarck's strategies, for example; wherever the political machinery of a state diffuses the electorate's division by numerous choices – wherever any or all of these complicating factors enter the national political scene, a multi-party system finds its *raison d'être*. Obviously, it reflects in its more numerous groupings a fuller and more exacting picture of the peculiar features of a stratified society than a two-party system could ever present.

Once established, these different party systems have far-reaching consequences for the voting process and even more so for governmental decision-making. So far as elections go, in a two-party system both contenders will naturally compete for the shifting voter in the middle. Politics, therefore, will gravitate towards the centre and the free-floating electoral bloc which is decisive for the winning or holding of a majority. The spectacle of a programmatic *rapprochement*, despite heated electoral campaigns, of the two main parties is not simply a trick of conniving politicians but a natural outcome of a party system that reflects a relatively stable order.

The double party has been called 'a convenient system for contented peoples' who are agreed upon the general principles of the constitution and the policies of the government and feel not too intensely about measures over which they disagree. The fundamental cleavage over the slave issue broke for a time the efficacy of the American two-party system. The double party structure of modern England presupposed the reconciliation of the British Tories with the House of Hanover. Once the Glorious Revolution with its political institutions was recognised, they in turn could absorb the shock and major changes of the following social revolutions within their unchallenged political framework. Thus the miracle of British democracy in its combination of stability and progress was guaranteed. Once effectively established as a political institution, the two-party system reinforces and often perpetuates the trend towards conformity.

A multi-party system does not possess this unifying and centralising order. On the contrary, yielding to its inability to produce a majority party in elections, its factions concentrate on the centrifugal forces of special-interest groups and may easily be directed at peripheral forces. Such diffusion of power, often emphasised by special electoral systems such as proportional representation (putting at high premium the true mirroring of all shades of public opinion), does not hold great promise for effective policy formation.

Policy-making in a two-party system is no doubt facilitated by its certain majorities and consequently its unmistakable party responsibilities. Yet even within the frame of the two major parties, a hidden multi-party system may often be detected with intra-party splits and factions or third parties, which may at times affect the crucial balance and enforce a break, new combinations or compromises at the political front. In these critical moments the direct mandate of the people may be blurred by the parliamentary struggle for majorities until the voice of the electorate can be heard again.

What is exceptional in a two-party setting, however, becomes the rule under the procedures of a pluri-party system. By its very nature, it must transfer crucial decisions from a much-divided electorate to the parliament. This representative body in turn must content itself with a government by the formation of a coalition, attaining, after extended bickering, the compromise which in the two-party system is largely the voter's business to achieve within his own mind and political wisdom.

Even such a sketchy outline makes it obvious how much the character and policies of parties are defined by this formal division between two-party and multi-party systems. A classification along this line, therefore, proves to be quite suggestive and essential.

The differences are even more fundamental in the comparison of the one-party rule with the two-party and multi-party organisations. The character of the totalitarian rule indicates that a one-party system is a contradiction in terms. For clarity's sake it might be better to assign to this important contemporary phenomenon a different name, such as 'political order'. Modern usage, however, has coined the concept of the one-party state, and, if only for reasons of propagandistic appeal, the terminology has been accepted on the daily political fronts, both national and international. This being the case, one might recognise in the opposition of the one party versus the two and multi-parties the fundamental cleavage of our time: *dictatorship versus democracy*. Along this dividing line all basic tenets of modern parties must be redefined: their structure and strategy, their leadership, their apparatus, their following, their techniques of mass communication, their national policies and their international ties. Through the organisation and practices of these contrasting party systems, in turn, the two great contenders over political control in our society stand fully revealed. The essence of democratic and dictatorial rule is embodied in the daily life of their parties.

Essential though this sharp hiatus is for a classification of modern parties, such a crude black-and-white comparison cannot suffice for a substantial analysis of the political processes in our time. It leaves too much unsaid and unexplained: the shaded differentiation between existing political orders along the line of democracy and dictatorship, the twilight zones of their correlations, the conditions under which a transition from one to the other is facilitated, the inner complexities of the functioning institutions in both systems and the shifting position occupied, in recent decades, by the modern parties in relation to the individual and the community at large.

The detailed analysis of dictatorial parties shows to what extent democracy is not only a symbol of aggression but also a claimed property of modern totalitarianism. The organisation and incorporation of the people become the overwhelming concern of the seemingly independent autocracies. On the other hand, the effective democratic party has also radically changed its character and role in modern society, although this fact may not yet be fully recognised by its adherents.

Perhaps it is altogether impossible to try to put into systematic categories such an intricate and constantly shifting political process – but, at best, one might conceive a set of 'emergent concepts' which would be highly significant. If one dared to measure at this point the long-range trends of party development, one might arrive at the following conclusions.

Modern parties have steadily enlarged their scope and power within the political community and have consequently changed their own functions and character. In place of a *party of individual representation*, our contemporary society increasingly shows a *party of social integration*.

This shift must be seen within the context of our changing society and its underlying philosophy. Three major stages can be observed in its development. Modern parties originated with the drive of a rising, self-conscious middle class that fought for liberation from the shackles of a feudal society and for representation to check monarchical absolutism. While the French Revolution officially proclaimed the end of this first phase of modern social development, the successful emancipation of rational man from the bonds

of the *ancien régime* and its caste system proved to be only a transitional second stage. The individual, set free, was soon striving at reintegration into a new society. In fact, since the middle of the nineteenth century diverse claims for such a new orientation have been raised, promising to stop the fragmentation of a *laissez-faire* society. The first and lasting challenge of rising socialism, the emergence and appeal of political irrationalism, and an awakening social liberalism gave contrasting answers to this key issue of our century. The dislocations caused by the sweeping industrialisation, radical urbanisation and international migration, by world wars and total revolutions, gave substance to a planned search for a new social order. We are still in the midst of this third phase. It constitutes the crisis of modern society.

It is against this background of crisis that a new concept of party is evolving. Its emergence and persistence, in fact, may well depend on the momentous character of social crisis. The well-balanced communities of the Scandinavian states and the Anglo-American world seem to be least affected by this new type, while it has found its most complete expression within nations in the grip of revolutions. The islands of social equilibrium, however, have shrunk, and the party of integration has no doubt become a salient feature of our contemporary landscape.

The *party of individual representation* is characteristic of a society with a restricted political domain and only a limited degree of participation. Its membership activity is, for all practical purposes, limited to balloting, and the party organisation (if existent at all) is dormant between election periods. Its main function is the selection of representatives, who, once chosen, are possessed of an absolutely 'free mandate' and are in every respect responsible only to their own consciences. This conception of an ephemeral party as a mere electoral committee does not correspond to the political reality and practice of the modern mass democracy, a fact which in many countries has been recognised (though often most reluctantly) in the crucial controversy over party discipline and even in numerous court decisions codifying party regulations, responsibilities and prerogatives. The fundamental concept of party, however, has hardly been challenged within democratic thinking.

Under the cover of such a persistent framework and rarely perceived even by circumspect political observers, a new type of party has emerged – the *party of integration*. The claim with which this party approaches its adherents is incomparably greater than that of the party of individual representation. It demands not only permanent dues-paying membership (which may be found to a smaller extent within the loose party of representation too) but, above all, an increasing influence over all spheres of the individual's daily life.

The first example of such a new party was presented by the Continental Socialists. Their organisation has been jokingly characterised as extending from the cradle to the grave, from the workers' infant-care association to the atheists' cremation society; yet such a description articulates the intrinsic difference from the liberal party of representation, with its principle of 'free recruitment' among a socially uncommitted, free-floating electorate (the bulk of which, in reality, may not be so independent). The following of the new movement is, indeed, much more clearly circumscribed by its permanent membership, the definite class alignment of its voting population and its far-flung participation in overall social affairs. The party can count on its adherents; it has taken over a good part of their social existence.

Despite such extensive organisation and intensified ties of its partisans, the Socialist party (and in an even more limited way the Catholic movement and other democratic parties of integration) include only a small active core among its wider circle of mere dues-paying members and its even greater number of mere voters. In fact, this differentiation is at the base of the much-disputed 'oligarchical' tendencies of modern mass parties which permit a relatively small group to decide the political fate of the disinterested and apathetic majority. Still, what is important is that the party in modern mass democracies has generally taken on an ever-increasing area of commitments and responsibilities assuring the individual's share in society and incorporating him into the community. This is no mere usurpation of power by the politicians but the natural consequence of the extension of the public domain and the constantly increasing governmental functions in a reintegrated twentieth-century society.

In this sense the phenomenon of the *party of democratic integration* has become a matter of record. This fact makes it the more imperative to recognise its basic variance from the *party of total integration*, which has found its prototype in bolshevism, fascism and National

Socialism. This all-inclusive party demands the citizen's unconditional surrender. It denies not only the relative freedom of choice among the voters and followers but also any possibility of coalition and compromise among parties. It can perceive nothing but total seizure and exercise of power, undisputed acceptance of the party line and monolithic rule. The rise of this absolutist police state decrees the end of democracy, of constitutionalism, of communal self-government, of Western man and his inalienable rights, of political parties.

This radical juxtaposition should forewarn the responsible student of modern mass society against the threat of party petrifaction, but such a mortal peril cannot be met simply by a denial of the extended functions of modern parties and of their radically changing character – for the choice is not between the absolute state and the absolute individual or between autocracy or anarchy, as the great simplifiers and political demagogues make us believe. On the contrary, constructive thinking must concentrate on the much more difficult and urgent task of devising political institutions that allow for a new adjustment between the integrated society and the free individual. It is within such a realistic delineation of the fundamental prerequisites, present-day responsibilities and necessary safeguards of a democratic society that the sociology of modern parties must be re-examined.

IV. SOCIOLOGY OF POLITICAL PARTIES

The study of the sociology of political parties has been completely dominated by Robert Michels's iron law of the oligarchical tendencies of social movements which he expounded with intense single-mindedness more than a generation ago in his *Political Parties*. While a standard work without doubt, it has shared the fate of so many classics – to be widely quoted and rarely read. Unfortunately, little empirical research has been done in the field of political sociology either in defining Michels's underlying assumptions or even in testing the validity of his thesis. In fact, it has become almost an undisputed axiom.

The first prerequisite for a more accurate appraisal of the intricate web of social relations within modern parties is an exacting theoretical framework. Without it, all the concrete investigations in the area will be meaningless, undirected and just so much ballast in the scholar's bark. At this point only a few crucial concepts, issues and hypotheses can be presented that may be suggestive for a proper orientation of future research. Key terms such as 'leaders' and 'followers', 'participation' and 'apparatus', need renewed clarification and a realistic differentiation in respect to the divergent political systems. Only an age corroded by the demagogical simplifiers of modern dictatorship could postulate the idea of a leaderless democracy. Only a naïve countermovement of overconfident democrats could overlook the weighty and often dangerous part played by aroused masses in present-day autocracies.

There are leaders and masses in democracies, just as in dictatorships. Leadership is a prominent theme of any political order. Yet every period, every society, every political system, establishes a different interrelation between leaders and followers. What is therefore important is to recognise clearly the divergent character, function, selection and mobility, which vary in accordance with the whole structure of the prevalent political order. This social conditioning of modern leadership serves as a key to an understanding of the difference between democracies and dictatorships in our day.

Although differing greatly in time and space, two main types of political leadership can be distinguished: *institutional* and *personal*. These are types roughly corresponding to the contrasting political systems, democracy and dictatorship. The pre-eminent elements of leadership in a democracy are institutional. This fact, of course, leads to a difference in democratic leadership according to varying institutional structures. The unlike character of the British prime minister, the French president of the council and the United States president are a reflection of different institutional set-ups of these democratic governments and their respective party systems. This institutional character defines the personal qualities of democratic authority, its functions, its rise to eminence, its continuation and its fluctuation and limitations in power. Deference, which is no doubt a basic desideratum of any leadership, derives largely from the leader's skills, the extent of his knowledge, judgement and foresight, his strength of conviction, his policies, his ability to attain social cohesion and to give direction to social forces. Balance of mind, distaste for violence of expression, a faculty for spreading conciliation on all sides, become highly valued

qualities of democratic leadership. One should add, to enlarge upon Bagehot's famed remark, that the leader must be 'an uncommon man of common opinions', which he articulates, intensifies, directs and realises through workable channels. This faculty of expressing the people's will points to a certain demagogical feature even in democratic leadership which is necessitated by the mass character of modern society. The pre-eminent elements of democratic authority, however, are institutional.

Dictatorial leadership is personal. Here again will be found variations due to national traditions, historical circumstances and personality patterns; but all modern dictators have this in common: they are anti-institutional. In fact, their very rise indicates the weakening or non-existence of political institutions, of a ruling class, of an accepted code of rules, of a belief in a rational order. The modern dictator is the substitute for institutions in the mass age of political confusion and social disintegration. He is, above all, the demagogue – the leader of the people, who, rising from the dark as an unknown soldier, breaks through society's institutional barriers. He does not stand for a positive programme but only for himself. He is responsible to no man but to God and the nation (who are conveniently removed from any direct interference), and in his very irresponsibility he is revered by the emotional, rootless and amorphous masses seeking mystery, devotion and the miraculous. He is the wonder-performing *charismatic* leader whose charm of personality conceals conflicts of policy. The 'state of the masses' is indeed the key to the acceptance of his myth.

Charismatic leadership and bureaucracy may well accompany each other. Even though it may seem inconsistent with the emotional dynamics of this personal ruler, his power rests on a strong party organisation, an apparatus. In fact, the guardians of his machine, the 'lieutenants', reveal the true character of the one-party system, its daily life, its divided rule, its inner tensions, its chance for survival and succession. Generally speaking, it is on this second level of political control that the multifarious functions of modern parties become apparent, the differences in degree and kind between the democratic and dictatorial parties are manifest and the resilience, flexibility and deep antagonisms of contemporary political organisations are tested.

Bureaucratisation is a categorical fate of modern society, involving equally government, business and political parties. This fact may be deplored for its social strains and its often serious consequences of hampering spontaneity and free play in human affairs; it is still an indisputable necessity inherent in an efficient, highly developed mass organisation. It implies stratification and hierarchy of office, specialisation and fragmentation of the citizen's role, centralisation of government. It has become 'the organisational weapon' of modern dictatorship. Yet its perils are not met by denying its existence or by warning of the inevitable destruction of the democratic fibre by the spider of administration. What must be stated carefully, however, are the manifold expressions it has found, the conditions under which this necessary phenomenon may choke the basic forces of a free society and the safeguards which can and must be erected against such constriction today and tomorrow.

As a preliminary investigation, an account should be given of the mulitfarious organisational forms within the diverse parties around the globe; possibly a comparison with developments in the adjacent trade unions, business organisations, pressure groups and social agencies would prove fruitful too. On the basis of these rich findings one might well arrive at some major qualifications of Robert Michels's glittering generalisations. They not only might be found to be an overstatement of the undeniable existence of oligarchical, hierarchic and centripetal trends in every society, but might be matched by data confirming the equally omnipresent democratic, levelling and centrifugal forces encountered even within present-day totalitarian orders.

What is more important, Michels's emphasis may reveal a deeply embedded bias in favour of the predominant bureaucratic, authoritarian values of his time. In this sense the initially socialist opponent of the Wilhelminic empire proves to be the latter's true child, who by no mere accident turns Fascist as a man. Such social conditioning may also be recognised in the crown witness of his inquiry – the German Social Democratic party – which, not without some justification, had been called one of the most Prussian institutions in the Reich. An analysis of other political movements growing out of a different social climate, such as the British Labour party, might have led Michels to very different conclusions.

Here is a random list of the multi-factored determinants within the internal party organisation which can help to define the

conditions under which its complex relationship of leaders and followers may take a line different from that postulated by Michels.

First of all, variations in the character of those who are led are as important as those among those who lead, especially as the *concept of the masses* presents an even more elastic term. Leaving aside frequent moral connotations (such as 'canaille', 'the great unwashed', the 'rebellious masses'), sharp distinctions should be made between the social variations of rural and urban groups, of scattered crowds and congregated mobs, of the latent and aroused masses, of the illiterate, numb and untried and the educated, alert, skilled people. Different social classes invite a variety of stimuli and reactions. Political parties find an uneven appeal in various social strata, in accordance with their specific social experiences and historical conditions. Le Bon's influential school of crowd psychology, generalising the experiences of the revolutionary upheavals in modern European capitals, identified the masses with the highly emotionalised and unstable character of the mob, thus stigmatising mass action as devoid of individual self-control. While his simplifications still have currency in popular thinking, highly sophisticated social research is beginning to uncover the much more complex interplay between the rational and irrational elements of modern mass reaction. Its findings will give a more adequate evaluation of mass participation in modern political movements.

Next to the specific character of the rank and file, the people's role in politics will depend on the *size of organisation*. The transition from aristocratic factions to modern mass parties no doubt necessitated the development of bureaucracies, as in any other large-scale organisations. All other factors being equal, this must diminish the individual member's potential in policy-making. Moreover, time, continuous effort and skills are increasingly demanded for adequate legislative decision and subsequent administration. Naturally, these prerequisites lead to the predominance of the permanent and party professional over the political amateur. The problem of political organisation, however, is not solved by abolishing the experts but by finding methods of checks and controls by an alert citizenry. Danger spots for a dynamic democracy appear where an incumbent party administration has a permanent hold on the organisational machine, absolute control over financial resources and monopoly of the internal channels of communication, thus stifling the voicing of counter-propaganda and alternative solutions and the rise of a substitute elite. The concrete situation in each nation and movement will depend on the complex of balances between party leadership and its apparatus, between parliamentary representation and national party conferences, between the executive core and peripheral membership.

Strength or weakness of the central authorities depends also on the *functions* to be performed. The single function of electioneering in the party of liberal representation will easily leave the power in the hands of 'notables', just as trade unions with an exclusive orientation towards labour contracts will not encourage wider membership participation. Not only do diffuse functions call for diverse leaders (efficient administrators and popular agitators, feudal confidants of the No. 1 man and reliable representatives of the interest groups, powerful masters of violence and political bosses, 'ambassadors of the people' and informal holders of effective party patronage), but their consequent contradictions and controversies also offer the adherents alternative choices and increased influences. Even, and especially, the seemingly monolithic rule of modern dictatorships in their claim to total controls had to engage a multiple leadership that caused persistent tensions among the lieutenants, the inefficiency of secretly competing hierarchies within the party, and final disintegration of arbitrary government under the stress of supreme crisis. The failure of fascism and National Socialism might well be ascribed to the lack of open competition, of circulation of elites and of free mass participation.

The *degree of participation* is a further key to the leader–followers relationship in political parties. Here again a preliminary exercise in definition is needed to clarify a politically charged and often purposely equivocal vocabulary. Ambiguity has become the secret weapon of modern autocrats. They will pride themselves on extending community enterprises as a genuine test of their 'people's democracy'. True, active involvement of a great number of party members can serve as a powerful counteracting force to oligarchical rule, yet democracy is not guaranteed by mere participation. If this means controlled and manipulated mass direction through the dictator's triple threat of all-embracing institutions, all-consuming fears, and an all-pervading propaganda, it can in

fact become the most potent instrument of his lasting power. Indeed, so far as his rule aspires to a radical change in the societal structure, effective mass participation seems to be a prerequisite for the successful breaking of the traditional patterns, for channelling the people's discontent and for preventing a counter-revolution. In this sense modern autocracies are still children of a democratic age; they are post-democratic dictatorships, and, because of that, they out-demonstrate any popular government in master-planned, minutely organised, mass meetings. Yet 100 per cent plebiscites are no demonstration of democracy, and true participation depends on more than 'bread and circuses'; it demands political activation through free choice and decision. In order to assure such a democratic process it must, above all, keep its channels of communication clear and accessible to those who are ready to attend to their citizens' rights and duties.

A realistic appraisal of functioning democracies must, of course, take account of variations in popular participation in accordance with innate abilities, acquired skills, available time and, above all, personal desires and aspirations. Negatively the degree of apathy is indeed a most powerful and measurable factor in the failure of effective democracies. It is not the least conditioned by the promise of rewards in the sharing of power, although a careful analysis will also have to consider differentiations in popular participation in democracies without simplifying and unrealistic assumptions of equal sharing for *all* citizens in *all* actions at *all* times. The role of the expert, of the administrator, of the responsible policy-maker, will need a fresh evaluation.

The morale of the group depends largely on the extent of participation of its members. The degree of *fluctuation* in leadership, as enunciated by Mosca and Pareto in the suggestive theory of the '*circulation of elites*', is a visible indication of the democratic chances of the party members. It gives a measure of the extent to which the controlling positions are open to free competition or monopolised by a closed group. It should also offer proof of the stability and longevity of the ruling class, although only isolated studies like Pontus Fahlbeck's classic on the Swedish aristocracy and similar works by Schott, Furlan and Savorgnan on Continental development have so far presented the kind of exact material necessary for comprehensive comparative analyses.

Equally, if not more, important for democratic vitality is the party's ability to absorb and assimilate new social strata into the ruling class. Closely connected with this key problem is the much neglected phenomenon of the *succession in generations*. Many political revolutions have originated in this natural break between different political ages; the continuity of a political system, on the other hand, can well be measured in terms of its ability to guarantee a smooth and steady transition. This test of political wisdom and viability becomes manifest in the diverse selection and training of future leaders (and the mechanisms by which leaders can be retired). These procedures offer the most conclusive clue to the aims, character and flexibility of the different party systems.

The usual *apprenticeship* of democratic aspirants through institutional channels of parliament, local government and party organisation is a selective process which tests particular qualifications, for framing legislation, for mastery in parliamentary debate, for teamwork, co-operation and successful compromise. It offers a severe training and a slow rise to power – too slow for some young men in a hurry, who join extra-parliamentary movements that seem to promise a short cut to victory. The institutional path attracts certain personality types who may be groomed preferably by a definite educational system and who are more easily found in specific professions. The availability of trained political participants and, to use Max Weber's terminology, of 'dispensable occupations' (lawyers, journalists, educators), which allow for career interruptions with even possible gains in skills and connections by such a political experience, will assure the filling of the necessary leadership reservoir. Equally important for the free flow of talent into political channels is a follow-up, a mechanism by which leaders can be retired and/or used for further advanced assignments without loss of craft and prestige. The rigidity of certain parties (and even more so, of many trade unions) is often due to their petrified leadership which has no exit but oblivion.

While it is undoubtedly true, as Lasswell has stated, that stability of rule varies inversely with the degree of organisation of a counter-elite, the inner corruption of a power group is equally accelerated by the absence of an alternative challenge. And since the weakness of authority is often more significant for the breakdown of a system than is the strength of the opposition, one may perceive some chance for the disintegration of a monolithic

rule, even if it has succeeded in destroying all articulate resistance.

Selection of leadership in modern autocracies meets with certain difficulties, despite the fact that conscious, long-range efforts are made for its preparation and for the master's succession. There is no natural competition in parliament or elsewhere to groom the dictator's lieutenants. Favouritism plays an important part in the selection, and especially in the higher brackets of the political hierarchy the leader's personal choice may even bring non-party members into the inner council. The main channel, however, is the party machine. Being a 'charter member' of the party (the old guards of the first hour, the martyrs of the preceding 'system's' prisons) is equivalent to a 'safe' seat in a democratic parliament. These important qualities of the ardent fighters and trustworthy confidants may, however, at times become embarrassing after the seizure of power, and the 'comrades of early and trying conspiracies' may be faced with purges and even the firing squad.

The composite structure of modern dictatorship in its all-embracing power requires varied types of leaders – demagogical, bureaucratic, militant and feudal – and the party machine has to put forth the necessary material. This often discordant variety of functions creates, in fact, a hidden multi-party system within the monolithic regime, bringing about selection through partisan factions, victorious before the unpredictable court of the supreme potentate, where the sycophant wins a friendly hearing and qualified potential is discarded. Such arbitrary selection may in the end cost the dictator his effective rule, especially in a time of great trial, as the demise of fascism and National Socialism amply illustrates. As long as he lasts, the modern tyrant has, however, the audacity to bid for the perpetuation of his regime even beyond his life by usurping complete control over the mind of the nation's youth. This monopoly of training through state-controlled youth movements and a completely co-ordinated educational system may be modern dictatorship's most potent weapon. It is certainly too early to draw any definite conclusion as to whether it will succeed where its classical forerunners failed; yet certain prerequisites can be recorded.

Two preconditions for successful *perpetuation* of the dictatorial system seem to be indispensable: (1) the establishment of a well-functioning party machine and (2) the destruction or absorption of the traditional institutions. Success or failure will depend largely on the degree to which the system is internally consolidated. If the shots fired at Lenin in August 1918 had killed him, the Soviet regime could easily have collapsed. When he did die in January 1924, an integrated power system had grown up around him, and, after the short struggle between Stalin and Trotsky, the master of the party machine emerged as the victor. Similarly, the 'heir apparent' of Stalin's regime had first been Kirov, and after his mysterious assassination Zhdanov took his place until he preceded his superior in death. If Khrushchev, following Malenkov, is destined to inherit Caesar's mantle, as has been widely assumed for a long time – the present transitional committee leadership notwithstanding – then the succession will adhere to the established procedures. The sequence of the list of 'No. 2' men in Nazi Germany from Göring to Hess to Bormann indicated that Hitler followed similar succession plans. Apparently, Goebbels or Himmler never had a chance at the 'line of succession' set down by Hitler. On the basis of our experience, which is extremely limited, we may surmise that the end of the political 'glamour boys' of the revolution and the rise of the less conspicuous middle-of-the-road *apparatchik* may well describe the bureaucratising pattern of dictatorial succession and – if it prove triumphant – the changing character of the revolutionary one-party regime.

V. THE PARTY IN A PLURALISTIC SOCIETY

This preliminary discussion of the sociology of political movements shows that a full definition of modern parties must include a clear understanding of the society of which they are a part and which they express. The nature and ethos of the ruling elite (military, managerial, intellectual, racial, religious, industrial, commercial, peasant or proletarian, as the case may be), the status differentiation and degree of mobility (class or caste structure), the prevalent value system, the inner cohesion or crisis character of a nation – the complex interplay of these heterogeneous factors can alone indicate the extent of oligarchical trends or democratic participation within a political community.

Even the one-party state, with its seeming monolithic control by its dominant clique, cannot escape the intricacies of competing agencies in a highly developed society if it wants to survive. The

purposely pluralistic democratic state must establish an ever-renewed balance of society's divergent forces in order to assign the parties their proper place and function.

In fact, a high degree of diversity seems to create a favourable climate for a functioning democracy if it protects itself against the perils of caste-like petrifaction and if it allows for a free interplay of its innumerable associations. And the varied parties serve as their forum.

Dictatorships, on the other hand, flourish among amorphous masses. If the natural groupings are not already pulverised, modern autocrats have to create a monolithic atmosphere, within which independent individuality and autonomous group life are submerged under the one-party rule and a 'classless society' decrees 'equality' before the tyrants. The fatal transformations in modern society – wars, inflation, depressions and revolutions – prepared the ground and recruited the crucial strata of a dispossessed middle class, a propertyless peasantry, a rootless unemployed, and pugnacious partisans and perennial soldiers of fortune. Such constitute the social raw material of the rising autocracy, which systematically dismantles all free agencies and reduces them to the level of enforced conformity. And this means the end of a functioning democratic party system.

The democratic process has to respect and integrate the numerous special interests within a live society. This creates the intricate and fascinating interrelations between *parties* and *pressure groups*, which are not identical and do not exclude each other; they coexist in continuous interdependence. Parties are not merely the sum of all pressure groups; nor can the public simply dismiss the interests' proper functions by muck-raking exposés of lobby activities. For a long time parties and pressure groups have had a bad press, largely because of a mistaken classical approach, which udged politics from the isolated individual's point of view. Once the social group is recognised as a proper starting point, the balance between the representation of homogeneous pressure groups (seeking influence) and the decision-making activities of heterogeneous parties (seeking office and reconciling the diverse forces within the state) becomes a major theme of national politics.

The techniques of interrelations differ in time and space. Pressure groups may try to establish or conquer parties, as has been the case with a number of European parties with strict class alignments or straight nationality backing. They may play the role of a free-floating vote bloc, seeking out the highest bidder, as practised in Anglo-American patterns; they may extend their influence through divided representation in different parties (*Querverbindungen*); or they may concentrate on minor parties, pressing their special interests.

Whatever their methods, the strength of the 'invisible government', as William Allen White called this consequential agglomeration of more than a hundred thousand associations of all types in America, should not be underestimated. Their power rests largely on their singleness of purpose and their great degree of centralisation. This directed appeal to specific concerns has evoked a most active response from the people, as the effective mass organisations of the trade unions and employers' organisations in the Western world testify, while a deep-seated scepticism concerning the political parties prevails. Unaccountable to the public at large, the economic pressure groups have often overshadowed the political parties and have, in fact, hindered them in their primary task of integrating the special interests into the framework of the national whole. A merely descriptive analysis of what is happening on the political scene is misleading. It seems to confirm the make-believe of politics as the sum total of pressure-group activities. Constructive leadership, while recognising their claims, must guard against their aggressions and, in the knowledge of their essential role, must lift the demands of the specific interests to the level of national needs. This is where effective parties come in.

This crucial problem of rational integration is not solved by the creation of 'economic parliaments'. On the contrary, useful though such bodies may be as supplementary advisory councils for national legislation, wherever they are introduced in place of political parliaments, they decree the end of democracy. The 'expert' character of occupational representatives limits their authority and naturally suggests the creation above them of a political apex which claims to represent the whole of the nation. Bismarck and Mussolini were not accidentally the champions of a chamber of corporations and the declared enemies of political parties. It is, indeed, through the network of political parties that the place and responsibilities of pressure groups must be circumscribed if the modern society is not to deteriorate into a neo-feudalism of powerful interest groups.

While the correlation of party and pressure groups decides the smooth flow of social forces, the political process is even more dependent on a proper delineation of *party and bureaucracy*. Here again the differences between democracy and dictatorship are obvious. It is equally true that the one-party state has in no way solved the dualism of the two forces; nor can democracies claim to have established an airtight distinction between the policy-making party government and the executive civil service. Theirs is a fluid relationship dependent on the continuous interplay of the agencies and, last but not least, on the incalculable informal contacts between the legislative and the executive branches.

The need for bureaucracy in the modern state is unquestionable, and so is the necessity for continuous checks on governmental agencies by the political parties in a democracy. How to combine the expert and the politician, administrative efficiency and democratic control, is a daily test of the ingenuity of political thinking. The British parliamentary rule of the responsible cabinet, the American presidential system and the continuity of the French permanent civil service (*vis-à-vis* constant cabinet shifts in a multi-party system) – all suggest different solutions.

Leaving aside popular slogans of 'the managerial revolution' and 'the dictatorship of bureaucracy', one must recognise that policy formulation in the highly complex and time-pressured present-day state is done largely by the administrative bureaux, while the functions of the political organs are limited primarily to the control, articulation and communication of public affairs. No simple formula can describe this intertwining process. Each nation has been trying in different fashions to adjust the tenuous relationship between bureaucracy and parties. Much spadework on this aspect of comparative administration is still needed to arrive at a fuller appreciation of possible solutions. To mention only a few outstanding problems: selection, training, ratings and reforms of civil service; the extent and limitations of patronage appointments; the right of bureaucrats to participate in party politics and to accept political offices, parliamentary mandates and cabinet positions; the actual influence of bureau chiefs and permanent under-secretaries in policy formation; the scope of delegated legislation; the administrative functions and control chances of political ministers; the changing attitudes of parties in power and in opposition to governmental agencies – all these much-discussed and controversial issues are indicative of the fact that it remains one of the major tasks of modern democracy to strike a new balance between executive and legislative power.

Confronted with such a challenge, the one-party regime finds itself in an insoluble conflict. The administration must be co-ordinated with the one and only political will and purged of all 'undesirable' elements in a much more comprehensive fashion than the natural shift of political appointees indicates in the democratic range of government. (This 'cleansing' from 'party-book appointees', cynically termed a 'restoration of civil service', is in fact a crude guise for securing promised rewards to 'deserving partisans'.) Yet the same dictators must show gratitude for the bureaucracy's service in bridging the old and the new order. The result is a discrepancy which reveals not only an inner party struggle between moderates and revolutionaries but the even more significant clash between state and party.

Dictatorial systems have varied in their attempts to meet these inherent contradictions. Fascist Italy alone went so far in at least posing the question of the self-dissolution of the Fascist party after its conquest of the state. This very suggestion may be proof of fascism's relative conservatism and state absolutism or of its utter failure to subdue Italy's bureaucracy.

Soviet Russia, on the other hand, decreed the complete abolition of the traditional bureaucracy; and the original Bolshevik theory even predicted the 'withering-away' of the state altogether. In reality, Lenin outstripped Peter the Great, and Stalin outdid Ivan the Terrible, in establishing the mammoth totalitarian apparatus in a super-planned society. The results were wavering methods, continuous fluctuations between the radical purge of a managerial class and the creation of a huge amount of red tape, suspicion against the 'experts' and unlimited power for the machine. The N.E.P., the calling of foreign specialists, the praise of the 'shock workers' (Stakhanovites), the temporary independence of the Red Army officers – all these measures, half-heartedly undertaken and repeatedly reversed, showed the indecision of the Soviet regime in respect to modern management and testified to the fatal failure of a party dictatorship to arrive at a *modus vivendi* with an efficient bureaucracy. Of course, the ruling party could also – and did, indeed – turn such a misfortune into an asset by making the

resistant bureaucracy a welcome scapegoat for all the ills of the monolithic order.

The Nazis' seeming success in establishing a monolithic order turned out to be an even greater failure. More revolutionary than fascism and yet more loaded with traditional liabilities than bolshevism, Hitler's Germany tried a weird solution of an artificial separation of state bureaucracy and party machine – a 'Dual State'. This separation between the party and the state (the one free from official handicaps expressing the 'boiling soul of the people', the other not responsible for this 'spontaneous outburst') could be clearly played up in an intricate technique of legalising revolutionary upheavals. It was a useful device to cover up irresponsible party action with 'respectable' formal legislation and thus to calm legal-minded people. The master-plan to enrol the 'neutral experts' in the strange combination of 'arbitrariness and efficiency based on order' was the essence of the Third Reich. But it did not prevent continuous clashes between the two hierarchies of party and bureaucracy. The discrepancies were fully revealed in the critical years of the Nazis' war on the world. After the party's collapse the bureaucracy still had to pay for the insincerities of the 'Dual State'.

Another critical relationship, especially in dictatorial systems, is found between *party and armed forces*. Democratic parties, by definition, exclude the creation of private armies, recognise the monopoly of military force in the hands of the State and exercise their power through ballots, not bullets. It is only in their resistance to modern aggressive movements that constitutional parties in Europe have half-heartedly considered the development of a paramilitary arm. These auxiliary forces, in fact, have not halted the attack against constitutional government but have only epitomised its loss in power and prestige.

The control of the armed forces is a question of survival for modern dictatorship. The military are usually the personification and last stronghold of traditional power, which must be crushed as the only serious threat to established totalitarianism. Moreover, their services are badly needed for a system directed towards war. This fact gives the professional soldier, especially in countries with a military tradition, a preferred position and a prestige far superior to that enjoyed in the 'unsoldierly bourgeois democracies'. In the beginning the promise of promotion to career-bound, ambitious young officers and the semblance of seeming independence quickly flatter them into submission. Experts are often politically blind and usually raise no objections to having the policy-makers pass on to them the watchword in return for their 'neutral services'.

In most dictatorships, however, it does not take long, especially for the controlling military circles, to realise that their fundamentally rational institution is worlds apart from the irrational politics of the modern *condottiere* and his irregulars who cannot be held by the hard-and-fast rules of army discipline. They take the world by surprise and manifest the unpredictable daring of the amateur and restless adventurer. Partisan guerrilla warfare and total terror are their weapons, not the carefully calculated and purposely limited campaigns of the professional high command. There is no compromise possible between these two conflicting concepts of strategy in violence. The continuous clashes between the Nazi stormtroopers and the Reichswehr, between the Fascist Blackshirts and the Italian army of Marshal Badoglio's brand, between the Soviet political commissars and Red Army commanders, between the Peronisti and the Argentine army, all testify to the persistent difficulties of the one-party state in mastering the professional soldier. Success or failure entirely determines the striking power of the dictatorial regime and its persistence, for its militant character is at the roots of the modern totalitarian party.

The militant structure of the dictatorial party is indeed the strikingly novel element in modern political organisation. From its beginnings it is, above all, a 'fighting league', with its 'shirt' movements to 'protect' the peaceful meetings of the rising party, its punitive excursions against recalcitrant opponents, its strict hierarchical order and party discipline, its actual fighting of a civil war. Politics becomes a battle of two irreconcilable opponents who are out to win and thereby to destroy the nation's or the classes' arch-enemy.

This permanent revolution has to go on even after the state is conquered – opposition groups must be seized, the party must be kept in warlike preparedness and must be regularly purged. All this militancy at home is only a prelude to the march on the world of the revolutionary party. Every revolution which claims to be more than a mere change of government and which pretends to create a new social order must also stir up neighbouring countries

and awaken in them a spirit of deep unrest. This quasi-religious 'missionarism' knows no frontiers and, with the services of world-shaking ideologies as a moving propaganda arm, tries to recruit soldiers of the new creed in an international civil war. This appeal brings into dramatic focus another element of the modern party – indeed its most critical feature – the international impact of political movements.

VI. THE INTERNATIONALS OF PARTIES

Political parties, when fully unfolded, operate in three concentric circles. They are based on personal loyalties; they manifest themselves through numerous groupings within a nation; and they reach out beyond state boundaries with their ideologies and their organised 'Internationals'. There is no Communist monopoly on world-wide planning, spectacularly and noisily though the Third International (Comintern or Cominform, as the case may be) has claimed a unique and unchallengeable position in this global battle of the minds. The colour scheme runs the whole gamut of the spectrum. In fact, the basic character of any party may well be determined by its international behaviour.

Little reliable material is available on these international ties and reactions; and the less is known, the more make-believe, rumour and mystery enter the speculations of the troubled citizens. Yet in our time – when the frontiers between domestic and international affairs are blurred, and revolutionary upheavals, destroying the stability of a prevailing social structure, are reaching beyond national compounds – political parties have become international forces that must be studied, especially where they shun the broad daylight in their submerged activities and, like an iceberg, allow only a small portion of their power bloc to appear on the global surface.

These world-wide disturbances may be registered and measured through a threefold approach: (1) through an evaluation of prevailing political ideologies; (2) through an analysis of the changing social structures; and (3) through an appraisal of partisan strategies on an international plane.

In order to make them useful yardsticks in the present political labyrinth, ideologies must be reconsidered as to their real meaning. Such an examination will find that the accustomed classifications of political movements in terms of conservatism, liberalism, socialism, communism, etc., are misleading, to say the least. Not only have innumerable combinations (such as national bolshevism, liberal conservatism, conservative socialism, the peoples' democracy) confused the political fronts, but the fundamental concepts themselves have shifted their original meaning under the impact of a radically changing society. One may even doubt the simple contrast between right- and left-wing movements. Originally 'radical' groups, such as the Parti Radical Socialiste, have been pushed to the right under the pressure of newly appearing leftist movements. At the same time frequent attempts at bridging right- and left-wing radical organisations, insignificant though they may be in view of long-range developments, point at similar psychological attitudes even among deadly political opponents.

The proper recognition of these difficulties should not lead to the easy dismissal of the concept of ideology altogether, as is frequently suggested by short-sighted 'realists' in politics. On the contrary, ideologies cannot be taken seriously enough in the vacuum of a spiritual crisis when the underlying unity of Christian European traditions is challenged by a counter-church, when uncompromising conflicts along petrified political fronts are threatening, and when the drives for fundamental reorientation are of primary importance in the final line-up of the peoples – at least, wherever they still remain free to choose.

Ideologies are the key to an understanding of the long-range strategy behind the day-by-day tactics of political movements. The means of international tactics are subservient to the ends of the axiomatic political religion. In such a world setting ideologies have become the most powerful weapon in international politics, and even their ambiguity has proved to be a potent factor in the political orientation or misdirection of a confused world society.

This confusion is to a large extent due to the radical transformations of the major social classes, which naturally have transcended national frontiers. Shifting party lines the world over cannot be understood without a full realisation of the crisis of the middle class. It represents as important a hiatus, especially in Western society, as formerly did the collapse of the feudal order which culminated in the French Revolution. The coinciding changes within the proletariat are equally remarkable, whether it takes a

turn towards bourgeois-mindedness or towards further class isolation.

The awakening of the peasantry, dormant for centuries, is another factor that may have far-reaching repercussions. The hesitant beginnings and sporadic appearances of a green International deserve careful watching. Last but not least, the rising crucial strata of all dimensions – the unemployed, the dispossessed, the militant irregulars – all have become international types and, as such, influence in some comparable fashion the political fronts in all nations.

In fact, the internationals of all colours have based their strategies on certain expectations of supranational class behaviour. The beginnings of an international Christian Democratic movement (whose support by adroit papal diplomacy and its popular arm, the Catholic Action, is difficult to appraise), the resurgence of world liberalism, neo-Fascist rumblings, the revival of the Socialist Second International, the gathering of latent forces among the erstwhile resistance movements, the first international party alignments at Strasbourg's Council of Europe – weak though all these attempts may still be, they can count on a certain response within different nations. These movements ought to be studied not only as potential powers of the future but also in their direct and indirect influence on national policy decisions at the present time. Political campaigns here and abroad have been deeply affected by their global connections. The strategy and outcome of the Italian elections of 1948 and 1952, the Papagos victory in Greece, the Saar vote, the United States presidential elections of 1952, the West German elections of 1953 and the British elections of May 1955 might have been very different, had it not been for their international implications.

Even bolshevism, despite its watertight isolation, cannot escape the influence from without, especially if it wants to affect this outside world. In this light, the transition from the Comintern to the Cominform may reflect a shrewd and realisic reappraisal of the changing dynamics of modern society. This new strategy, while still insisting on Lenin's fundamental formula of class loyalty above national loyalty, seeks more than a mere proletarian base of recruitment and mobilises instead – if only for short-run tactical purposes – all elements of unrest within the nations. Now the party appeals, above all, to powerful national ambitions. It is not altogether a new turn. Mao Tse-tung and Ulbricht may take their line from Borodin and Radek of a quarter of a century ago. Revolutionary activists of the war resistance, an awakened peasantry in the shadow of feudal south-eastern Europe, China and south-east Asia, and a deeply disillusioned middle class everywhere can be steered into a new 'Fatherland Front' and might be united by the militant appeal of a grass-roots, partisan nationalism, by the fear of American 'imperialism' and by the promise of 'lasting peace'. Indeed, peace becomes a most potent political weapon of Cominform propaganda in a war-weary world. But, while the new Communist combine of the U.S.S.R. with its ring of satellites, as compared with the old Comintern, seemingly allows for greater national autonomy and a more decentralised system of Red internationalism, its main purpose still remains the establishment of a strong arm of Soviet foreign policy. Wherever the new nationalism comes into conflict with Russian interests, it must be denounced in the name of the revolution as a nationalist deviation.

The world revolution has found a fatherland. It gives the Third International a tremendous striking power, which a revolutionary state does not hesitate to exploit unscrupulously. This very connection, however, creates all the liabilities under which communism as the international arm of Russian bolshevism suffers within every nation. The divided loyalty of its members has meant continuous setbacks to the Communist parties within other nations, purges among their leadership and confusion for the following.

Red imperialism, by no mere accident, experiences increasing tensions within its own camp. Titoism is more than a peripheral phenomenon of an overexpanded super-power, although the fact of geography is not a minor item in this first defection from the Bolshevist creed. Geography may well be counted a deterrent to an extended defiance of Soviet rule among adjacent satellites which can be easily held under the sway of the Red Army and of militant bolshevism. The 'good neighbours' must present governments 'friendly to the Soviet regime'.

Indeed, in this age of *international civil war* – cold, lukewarm or hot as it may be – between the super-power blocs of this bipolar world, political parties are aligned in the name of, or in opposition to, *revolution*. Commonly regarded as a mere internal upheaval. revolution has become a world phenomenon. Its significance is measured by its international effect. Radical upheavals, as all

great revolutions are, must be played on an international stage. Every region has become sensitive to the developments of far-distant lands.

The very fact that areas far apart geographically and historically have been pushed together on this shrinking planet has created a revolution altogether different from the stereotypes found in story-books and even in more learned texts. The complexity of this new cataclysm must be fully comprehended in order to appraise the specific nature and aims, the spread and directions, of the contemporary upheaval.

Different ages and different revolutions have become strange contemporaries. They cannot be isolated in neat separate compartments – they fight, they influence, they imitate one another. The contemporary revolutions are inextricably mixed in structure and strategy, in patterns and politics. Still every era has its predominant type; the stronger revolution sets the style.

What is the character of the twentieth-century upheaval? It is neither an objective phenomenon outside human control, such as the interference of God or nature (in the definition of the Renaissance) within world affairs, nor is it the free, subjective human act of heroes standing up against and reversing the natural trend of objective phenomena, as the age of the barricades and the Romantic revolutions saw it. Now the 'objective' march of events and the 'subjective' deeds of heroic man are interwoven into a new revolution when 'the time is ripe' for it and when dynamic leaders can articulate the 'revolutionary situation' which otherwise may pass by. There is nothing accidental in its outbreak, no chance in its unfolding, no sudden whim in its success or failure. The modern revolution is a calculated, planned, long-range process and is executed almost to cold scientific exactness by the 'professional revolutionaries'. No longer can it trust to the appearance of the dynamic dilettante whom the spirit moves to enter meteor-like and to disappear when the excitement of the heroic battle is over. It summons the sober rationality of the patient expert. Organised political parties play a major role in preparing and continuing the permanent revolution. Like modern warfare, its outcome is often decided long before the actual declaration of hostilities. In fact the carefully planned revolution may well be over but for the shouting at the time of the official outbreak, as the Czech coup of February 1948 proved so tragically. Moreover, not a single soldier may be needed to move across national frontiers in order to win the major battles in this international civil war. A central revolutionary authority (under whatever name – Politbureau or Central Committee) can direct its orders across frontiers by remote control through the well-established revolutionary pipelines of the disciplined party and its organisational weapons.

Such an upheaval can be defeated by its opponents only if they recognise the changed character of modern revolutions and thus meet the long-range strategies of the Moscow manipulators by equally circumspect planning against the root causes of unrest – in time and with vigour.

The second feature of modern revolution – its confluence of different types of upheavals – is a case in question. Probably the most blatant example of this complex phenomenon is presented in south-east Asia and the Middle East today, where the democratic, the national and the social revolutions coincide. While the first and the second kind have been the characteristic revolutionary expression of the great world in the nineteenth century, they are now finally evolving in the erstwhile colonial areas. Yet their belated appearance in the Far East, south-east Asia, the Near East and south-eastern Europe makes them contemporaries of the twentieth-century social revolution. And as in earlier upheavals a stronger nationalism had penetrated and conquered the weaker liberal forces, today the dynamic social revolutions easily take command of the budding democratic-national movements and direct them into the stream of the 'coming world revolution'. This may not be surprising in view of the difficult plight of these newcomers among nations (the effect of long-lasting foreign political and economic domination, of 'imported' revolutions, of inexperienced leadership and of a weak middle-class basis). Such a turn is quickly offered by the master-strategists of the Soviet revolution, who, in accordance with the Leninist–Stalinist theory of imperialism, attempt to co-ordinate both social revolt and colonial liberation and therewith raise the claim to revolutionary world leadership.

Marxism, which within a highly industrialised society has met with remarkable resistance among the 'internal proletariat', to use Arnold Toynbee's suggestive phrase, may be more successful with the 'external proletariat' of the poverty-stricken masses of agrarian feudalism in underdeveloped areas. (Incidentally, it is significant that Italy and France, least dynamic in their industrial develop-

ment and still influenced by feudal traditions, have been the only Western powers where Soviet infiltration has made persistent and deep inroads). Imperialism, until recently a mighty weapon in the hands of distant exploiters, may be turned by rising colonial peoples into an easy scapegoat for all the ills that have befallen them under imperial occupation. The dangerously vacillating game of passing demagogues, like Iran's Mossadegh or Egypt's Neguib and their fanatical 'party gangs', must be seen against this sordid background of desperately impoverished economies. No doubt, revolutionary Marxism has been not without success in these areas of awakening nationalism, and this despite the fact that the drive for Westernisation has been the clarion call for the 'underdeveloped peoples', the U.S.S.R. included. The Point Four programme is an attempt at mobilising and recapturing the imagination of the colonial world and no less that of the United States, which, we must remember, was also born of revolution.

Late may be the hour in this great competition for the support of the new protagonists just passing the threshold of their conscious history. Time is of the essence; yet it may not be ours. 'The meeting of the East and West' has been a head-on collision brought about by wars and revolutions. This naturally means cultural dislocations and social tensions. To transplant complex production methods into a primitive economy and to telescope centuries of slow European growth into a short span of one generation presages serious perplexities. To direct the highly excited and equally immature masses will demand responsible leadership. Yet this individual sense of duty to the greater community may well be strange to oriental traditions. Besides, a germane political elite – one which has grown up in rebellion against imperial rule – may not be ready to take on the unaccustomed and different functions of governing. A new generation of leaders will be needed and may be wanting.

The development of a responsible party system could well be the secret of a successful transition from colonialism to political self-rule. India and Pakistan, despite their political 'childhood diseases', show certain promising signs. The relatively smooth surrender on the part of the British, especially in the last critical stages, may have something to do with it.

These are but a few of the burning issues faced by the awakening nations and no less by their aged contemporaries. The peoples meet, clash and conciliate their interests in the moves of organised parties on the international plane everywhere. They are often not sure of their own stand. Prelude or postlude, the intricate choices before them reflect the confluence of different streams of revolution and their consequent confusion. Hesitation and doubt also reflect widespread uncertainties as to the true nature and direction of the revolution in our time. Above all, however, this revolution challenges our ingenuity to articulate workable programmes, to organise functioning movements and to put them to constructive action – weighty responsibilities which rest primarily with the peoples' great intermediaries: the political parties.

Notes

1. S. M. Beer, in Sigmund Neumann (ed.), *Modern Political Parties* (Chicago, 1956), pp. 9–57.
2. E. E. Schattschneider, ibid.

A COMPARATIVE STUDY OF INTEREST GROUPS AND THE POLITICAL PROCESS

Gabriel A. Almond

THE first research planning session of the Committee on Comparative Politics of the Social Science Research Council was held on 5–10 April 1957 at the Center for Advanced Study of the Behavior Sciences at Stanford, California. The participants included some of the recipients of S.S.R.C. grants for field studies of political groups, as well as a number of other scholars planning field research on these problems. The purpose of the Committee in sponsoring planning sessions among its grantees and other interested scholars is to enhance the cumulative value of research efforts now under way or planned for the near future. As a result of the S.S.R.C. programme, as well as of a number of other organised and individual efforts, we can anticipate in a few years an extensive monographic literature dealing with political groups and processes in a great many foreign countries and a variety of different culture areas. Systematic information on this scale may not only fill in 'areas of ignorance', but offers an opportunity for significant advances in the general theory of politics. Our present theories on interest groups, political parties and public opinion are based on American and to a lesser extent European experience. We can now look forward to a situation in the near future in which

theories of interest groups, parties and public opinion can be built up on the basis of a rich collection of intensive studies.

The likelihood of a significant theoretical product resulting from this fairly large group of individual efforts may be enhanced by the discussion of hypotheses and research approaches prior to field work, as well as by the pooling and exchanging of information and hypotheses during the course of research and the subsequent analysis. The purpose of the five-day planning session at Palo Alto was to explore the possibilities of agreement about research objectives and methods, and to codify the conclusions of the participants in the form of a research programme statement. The planning session proceeded by plenary discussions and the presentation of specific reports by work groups concerned with special problems. The conclusions reached in the general discussions and the reports of the work groups are summarised in the present statement. The topics discussed during the session were as follows: (1) the general objectives of the comparative study of interest groups; (2) general description of the interest-group system; (3) the selection of specific interest groups; (4) interest groups and public opinion; (5) interest groups and political parties; (6) interest groups and the legislature; (7) interest groups and the bureaucracy; and (8) the inclusion of non-Western countries in the comparative study.

I. THE GENERAL OBJECTIVES

Comparative analyses of political institutions have thus far been confined to formal governmental institutions, and to political party and electoral systems. Dissatisfaction with these formal comparisons is widespread in view of the generally appreciated fact that formally similar governmental and party systems often function in radically different ways. And the search for explanation of the formally similar but differently functioning political systems has turned to vague residual categories such as 'social structure', 'national character', 'consensus' or its absence and 'public opinion'. In Duverger's study of political parties he concluded:

> The development of parties has burst the bonds of the old political categories inspired by Aristotle and Montesquieu. The classic contrast between parliamentary, presidential and National Convention regimes can henceforth no longer serve as the pivot for modern constitutional law. Kemalist Turkey, Soviet Russia and Hitler Germany were profoundly similar because each was a single-party state, although the first practised the National Convention regime, the second a semi-parliamentary regime and the third a semi-presidential regime. In spite of their common attachment to the parliamentary regime, Great Britain and the Dominions, under a two-party system, are profoundly dissimilar from Continental countries under a multi-partist system, and in certain respects are much closer to the United States in spite of its presidential regime. In fact the distinction between single-party, two-party and multi-party systems tends to become the fundamental mode of classifying contemporary regimes.[1]

In Duverger's subsequent remarks it becomes clear that even his distinctions between single-party, two-party and multi-party systems are inadequate as principles of explanation of the functioning of political systems. He and other scholars have noted the fact that the multi-party systems of France and Italy differ from those of the Scandinavian and Low Countries, and that the one-party system of Spain is rather different from those of Russia and Nazi Germany. To understand the differences between the multi-party systems of France and Italy and those of the Scandinavian countries we would have to look 'below' the levels of the party systems into the 'interest-group' systems of these countries and into basic popular attitudes towards political authority, partisanship and interest – their 'political cultures' in other words. The differences between the one-party regimes of countries such as the Soviet Union and Nazi Germany on the one hand and Spain on the other are to be accounted for largely by the different goals and practices of the groups controlling the authoritative governmental institutions.

In our search for types of polities and the principles governing their behaviour it would be a mistake simply to go one or two steps beyond Duverger, and say not the party system, but the interest groups and political culture '. . . become the fundamental mode of classifying contemporary regimes'.[2] The mistake arises out of the search for some single 'crucial institution' or principle of explanation. In other words we turn to the comparative study of

interest groups not with the hope that these rather than parties or governmental institutions will yield *the principles* of discrimination between types of political systems, but rather with the expectation that the systematic examination of interest groups in their complex interrelations with public opinion, political parties and formal governmental institutions will enable us to differentiate more accurately between political systems as *wholes*. In other words, the growing concern among scholars with interest groups and public opinion is the consequence of a search for a more complete and systematic conception of the political process as a whole, rather than a search for an approach which is an *alternative* to the present emphasis on formal governmental institutions.

As Truman and other scholars have pointed out, interest groups may be articulate or inarticulate, manifest or latent, formally organised or simply a condition of like-mindedness and informal communication about issues.[3] A comparison of Western and non-Western political systems brings out sharply the 'latency' of the typical non-Western political system in contrast to the overtness of Western politics. How different are the problems of public policy-making in a society where there are no effective mechanisms for the articulation of political demands, from one in which there is an elaborate set of structures for the formulation of demands, and the transmission of these demands into the party and governmental systems? The kinds of interest groups which are present in a society, the specificity or diffuseness of their demands, their conceptions of the political arena and of the 'rules of the game', the ethos which they bring with them into the political process: these are the 'raw materials' of politics – the unaggregated demands – which some set of mechanisms must transform into political personnel and public policy.

These general observations about interest groups not only suggest their importance as a subject of study, but set certain specifications in research design if the maximum value of a comparative study is to be attained. A good research job on interest groups in a particular country which may make possible meaningful comparisons with other countries, must examine the interest-group system in its relations with the social structure and culture on the one hand and the other parts of the political structure on the other. In identifying the interest-group system in any particular country this broad functional approach will prevent us from identifying interest groups with any particular kind of structure. The function of articulating and transmitting group interests may be performed in one system typically by the well-organised and bureaucratised 'pressure groups' familiar in the West, or it may be performed in another system typically through an informal and intermittent process of communication between and among class and status groups such as large landholders or businessmen, and cliques of bureaucrats and/or army officers. If it is possible to state the theme of the comparative study in the form of a single question it might be: What form does the articulation of political interests take in various societies, and how are these interests transmitted to other parts of the political and governmental structure, and translated into choices of political personnel and public policy?

II. THE GENERAL DESCRIPTION OF THE INTEREST-GROUP SYSTEM

It was the general conclusion of the planning session that each of the participants in the comparative study would attempt a general survey of the system of interest groups in his country, describing the kinds of interest groups which are present, their membership size in relation to potential, their financial resources, activities, patterns of coalition and interaction among them, and internal patterns of organisation and decision-making. The usefulness of such a descriptive census was defended on the ground that many of the most significant differences among interest-group systems will come out sharply through such description. Thus, three of the distinguishing characteristics of American interest groups – the large number of civic and ethnic interest groups, and the relatively high degree of membership participation in decision-making within interest groups – would come out clearly in such a comparative description. A comparison of the structure, the membership and the financial resources of the trade unions on the European continent will tell us a great deal about the differences between the political systems of France and Italy as compared with Germany, the United Kingdom, the Low Countries and the Scandinavian countries. The predominance of functionally specific, bureaucratised, associated types of interest groups in the West, and of kinship groups, status groups and of informal cliques in the non-Western areas is one of the most crucial distinctions

between these types of political systems. Again the clearer distinction between the functions of parties and interest groups in the United States, the United Kingdom and the Commonwealth, than obtains on the European continent and the non-Western areas, is a major clue to the special properties of Anglo-American politics.

The planning session was in full agreement on the desirability of such a general survey and appraisal of the interest-group system as a whole. At the same time the discussion pointed up clearly the radically different conditions of research in the areas represented which might make such a general survey a full-time problem in one area, and an easier task in another. For example, in a country such as Germany, the high degree of formal organisation of interest groups, the availability of directories, of interest-group publications and of monographic studies by German scholars, would make such a survey a relatively simple problem and quite consistent with a major allocation of research time to the characteristics and functioning of a limited number of selected groups. In India on the other hand even an attempt at an accurate survey of the kinds and characteristics of interest groups might very well constitute a full-time research operation. Even on the European continent the search for comprehensiveness of coverage and accuracy of detail in some countries might seriously conflict with other research purposes.

A review of the range of problems which might confront the participants in the comparative study in making such general surveys and appraisals led the group to the conclusion that these would in every case be ranked as a research goal, secondary to the more intensive analyses of the characteristics of specific, selected interest groups. They would be carried on primarily as a by-product of the documentary research and interviewing programmes of the participants. It was recognised that this would make some of these general appraisals less well documented and more speculative than others. It was also agreed that the general descriptive material gained from such surveys, and the general hypotheses about the characteristics of interest-group systems as wholes, would be reported by each participant in the form of a general introductory chapter or chapters in his monographic report. It was also agreed that these general appraisal chapters would be exchanged by the participants as early as possible in order to have included in the individual country studies propositions about similarities and differences between the interest groups of the particular country studied and those of the other countries.

III. THE SELECTION OF SPECIFIC INTEREST GROUPS

While all the participants in the planning session had expressed the intention of studying interest groups in their particular countries, considerable variation emerged in the kind of group or groups which had been selected for emphasis. The original preferences were based on prior research interests and the salience of particular kinds of groups in the countries studied. The planning session brought out the point that while the choice of group or groups differed from one country to the next, all the studies had in common a concern with the political function of articulating and transmitting demands into the political process. The different choices of types of groups was largely due to the fact that different kinds of interest groups performed these functions in the various countries. Thus in the Western European countries, the job of translating the interests of different segments of the society into political demands, and bringing these demands to bear in the policy-making process, was largely performed by the trade union movements, the agricultural federations, business groups and trade associations and church groups of one kind or another. On the other hand, in Asia, the Middle East and Latin America two major classes of interest groups were in operation, one typical of the more modern, industrialised and Westernised sectors of these societies, and the other typical of that part of the society less affected by change. Hence, a study of interest groups in the non-Western countries would have to sample both the older and the newer types of structures. It would have to look into the activities of the emerging trade union, business and student movements in the modernised, primarily urban parts of these societies, and the status groups, kinship and lineage, religious and ethnic groups of the village and the countryside. The problem of interest-group research in these areas is not only that of describing the characteristics and functioning of these two types of interest-group systems, but also examining their relations with one another.

In the course of the discussions it was agreed that, while each participant might give greater emphasis to research on a particular class of groups, some attention would be given to the activities

of the other significant groups. The argument here was that a study which, for example, concentrated on the political activities of trade unions would be incomplete without some attention to business groups and trade associations, agricultural organisations and religious groups. Hence, those concerned with the European countries agreed that some attention would be paid to each of the major classes of interests. Specific decisions about the allocation of emphasis and methods would have to be made by the researcher in the field, in the light of the availability of information, the relative importance of these groups and his special interests. Those concerned with the non-Western areas agreed on the desirability of sampling the 'modern, urban' type of interest groups, the interests operating within governmental agencies including the army and the traditional groupings. Here again the specific selections would have to take into account the special characteristics of the political system in the area, as well as the conditions of research.

The general conclusion was that by being concerned with the 'function' of interest in the political process, rather than certain kinds of structures, and by being concerned with sampling the interest-group universe, a significant step towards comparability and cumulativeness of theory would be accomplished.

IV. INTEREST GROUPS AND PUBLIC OPINION

One of the most challenging opportunities presented by the comparative study was that of systematically introducing public opinion data into the study of political interests. One of the central problems in interest group theory is the relation between manifest and latent interest. To what extent can organised, overt interests be taken as reflecting the interest tendencies of the general population? The phenomena of the mob in non-Western countries, of riots in totalitarian countries, of 'Caesarism', 'Poujadism' and 'incivisme' in the European area suggest that popular attitudes and tendencies are a separable factor in the political process, the properties of which cannot be inferred from the existing organised tendencies and from electoral behaviour. Any characterisation of a political system would be incomplete if it was confined solely to a description of current organisational patterns and processes. Latent interest may not only result in future changes in organisation and process; it establishes an atmosphere which affects the contemporary operations of the political process.

It was agreed that the relations between interest groups and public opinion would be an important dimension of the comparative study. One of the main, but by no means the only, source of information about public opinion would be the accumulations of public opinion data available in many of the European countries. In the course of the discussions of the kinds of polling material which were available it became clear that while there was much that might be useful, there was relatively little that dealt directly with the problems of legitimacy and consensus, that is, popular attitudes towards governmental authority, and the attitudes of groups in the population towards one another. Thus, certain basic hypotheses about the relations between the condition of political opinion and the functioning of political systems in the European area had never been directly investigated, but were now in the form of inferences from the behaviour of groups and institutions. The question put before the session was: Would it be possible through a survey of comparative political opinion to establish (1) that the British political process with its moderate interest groups, and its effectively functioning two-party system, is related to a homogeneous, secular and loyal public opinion; (2) that the French and Italian political processes with their fragmented party and interest-group systems, are related to fragmented, only partly secularised and largely alienated political cultures; and (3) that the political systems of the Scandinavian and Low Countries with their 'working multi-party systems' and their relatively moderate interest-group systems were related to political communities in which the 'consensual bond' was stronger than in France and Italy, and in which the great majority of the population viewed governmental authority as having a useful and constructive function in relation to their interests?

All the participants agreed to examine the accumulations of public opinion data in their countries and to use whatever information they could get from these sources on attitudes towards interest, party and state. In addition they agreed to consult and advise in the planning and execution of a comparative opinion survey which might deal more directly with these phenomena, if such a survey should be undertaken as a part of the programme.

A second aspect of the relations between interest groups and

public opnion was discussed: the propaganda and information activities of interest groups. The ways in which interest groups conceive of their audiences and the ways in which they represent their interests to the public should throw light on the functioning of the political system as a whole. For example, French business associations are different from the American in that they do not engage openly and on a large scale in public 'informational' activities. This may reflect a general condition of fragmentation in political communication in France, a condition of distrust and alienation among interests. The general proposition was advanced that certain patterns of interest communications activity are associated with certain basic conditions of the political community.

It was agreed that the 'public relations' of interest groups would be examined with the following questions in mind: (1) How does the interest group conceive of its audience? (2) How is the public relations function organised and what is the magnitude of the public relations effort? (3) How are interest groups related to the specialised communications system of the society (e.g. press, radio, educational system and the like)?

V. INTEREST GROUPS AND POLITICAL PARTIES[4]

In the discussions and conferences of the Committee on Comparative Politics over the last three years, and in papers prepared by Committee members and other participants in the Committee's work, the broad outlines of a theory of comparative political systems has slowly developed.[5] In part, this theory has been concerned with the development of models of the various forms of the political process in existence today. During the planning sessions in Palo Alto these models were again discussed; they influenced much of the research design which is reported below. It may be useful to summarise these characterisations of the Anglo-American systems, the varieties of non-Western political systems, the French–Italian 'crisis' system and the 'working multiparty' systems of the Scandinavian and the Low Countries.

In the Anglo-American type of political system the functions of political parties and interest groups are sharply differentiated. Interest groups articulate political demands in the society, seek support for these demands among other groups by advocacy and bargaining and attempt to transform these demands into authoritative public policy by influencing the choice of political personnel, and the various processes of public policy-making and enforcement. Political parties tend to be free of ideological rigidity, and are aggregative, i.e. seek to form the largest possible interest-group coalitions by offering acceptable choices of political personnel and public policy. Both the interest-group systems and the party systems are differentiated, bureaucratised and autonomous. Each unit in the party and interest-group systems comes into the 'market', so to speak, with an adjustive bargaining ethos. Furthermore, the party system stands between the interest-group system and the authoritative policy-making agencies and screens them from the particularistic and disintegrative impact of special interests. The party system aggregates interests and transforms them into a relatively small number of alternative general policies. Thus this set of relationships between the party system and the interest-group system enables choice among general policies to take place in the legislature, and assures that the bureaucracy will tend to function as a neutral instrument of the political agencies.

We might take as our second type a model summarising the properties of the political systems which are to be found in Asia, the Middle East and Latin America in which neither parties nor interest groups are fully differentiated. Associational interest groups such as trade unions and business associations may exist in the urban Westernised parts of the society, but in the village and the countryside interest organisation takes the form of lineage, caste, status, class and religious groups, which transmit political demands to the other parts of the political structure by means of informal communication. In one version of this class of systems parties tend to be *ad hoc* coalitions without permanent bureaucracies, and without grass-roots organisation. They exist primarily in election periods and in effect cease to exist in the intervals between. Given such weak and non-aggregative party systems the capacity of the legislatures to formulate alternative policy choices may be seriously impaired, as is their capacity to control the bureaucracies. In many of these political systems the significant political groups are neither the parties, nor the associational interest groups, but elements or cliques within the bureaucracy, and the army; and cliques, informal groupings and powerful families formed within such non-associational interests as religious communities, the large landowners, the business community and

the like. The political process consists of the informal communication and flow of influence between these informally organised interests, and groups within the bureaucracy and the army.

The instabilities of this type of political system arise out of the fact that the agencies for the articulation, communication and aggregation of interests are incomplete and unrepresentative, as well as out of the fact that the demands transmitted into the political system from interest groups are vague, diffuse and of radically unlike content and intensity. Latent interests, lacking overt and organised channels of expression, may suddenly break into the political arena. The information available to influential groups and individuals about the expectations and attitudes of the various interests in the society cannot be complete or accurate. Hence, calculation is impossible, and the flow of political interaction involves under-reaction and over-reaction, violence and apathy, alternations of periods of political latency, with sudden and violent shifts in power.

Given the basic instability of this general class of political systems, authoritarian stabilisations are a frequent event. Indeed, in many of them the developmental pattern is one of a shift from an unstable pluralism to authoritarianism, and then back again, or a shift from the authoritarianism of one clique to that of another. Authoritarianism may be based on control of the army either by a clique of army officers, or a clique of bureaucrats controlling the army, or by a coalition of both. Still another pattern is one in which the desire on the part of a controlling group to secure its own power and destroy opposition, or to mobilise the society for industrialisation and national expansion, leads to the formation of an authoritarian party which actually penetrates the countryside. In some cases as in Turkey and in India the objectives of the ruling groups and of the dominant party are tutelary. That is, the function of the party is not only control and mobilisation, but also political acculturation, the preparation of the ground for the emergence of a Western-type associational system and of a Western-type party system with a coherent, responsible and loyal opposition.

Thus, it should be quite clear that there are many kinds of non-Western political systems. They all appear to have in common (1) a fragmented political culture as a consequence of Westernisation, in many cases added on to an indigenous cultural heterogeneity; (2) poor political communications and a high degree of interest latency which renders political calculation difficult if not impossible; and (3) a party system which is incapable of aggregating and synthesising interest demands into a small number of political alternatives either of personnel or of public policy. On a scale of political differentiation one would have to say that certain kinds of structures such as associational interest groups, the mass media of communication and the kind of party system common in the West and essential for the functioning of a modern mass-suffrage parliamentary system, are present at best in only a limited degree. On a scale of functional specialisation one would have to say that in the absence of fully developed associational interest groups, party systems and modern media of communication, the functions of interest articulation, aggregation, communication and transmission are largely performed by bureaucratic or army cliques, traditional structures such as lineage or status groupings and by mobs, street demonstrations and the like, which serve as one of the agencies by means of which latent interests are articulated and transmitted.

A third type of political system is exemplified by France and Italy and by the Germany of the Weimar Republic. Contemporary Germany appears to be moving in the direction of an autonomous interest-group system and an aggregative two-party system – towards the Anglo-American model, in other words. In the French and Italian political systems parties and interest groups are organised and bureaucratised, but they are not autonomous systems. They interpenetrate one another and consequently fail to realise the two-stage pattern of the political process characteristic of the English and American systems. There are some parties which more or less control interest groups (e.g. the Communist party and the Communist-dominated trade unions, and to a lesser extent the Socialist parties and the Socialist trade unions). There are some interest groups which more or less control other interest groups and parties (e.g. the Church, the Catholic trade unions and the Catholic parties, business interest groups and the centre and right-wing parties, and the like).

When parties control interest groups they inhibit the capacity of interest groups to formulate pragmatic specific demands; they impart a political-ideological content to interest-group activity. When interest groups control parties they inhibit the capacity of the party to combine specific interests into programmes with

wider appeal. What reaches the legislative process from the interest groups and through the political parties thus are the 'raw', unaggregated demands of specific interests, or the diffuse, uncompromising or revolutionary tendencies of the Church and the movements of the extreme right or left. Since no interest group is large enough to have a majority, and the party system cannot aggregate different interests into a stable majority and a coherent opposition, the electoral and legislative processes fail to provide alternative, effective choices. The result is a legislature penetrated by relatively narrow interests and uncompromising ideological tendencies, a legislature which can be used as an arena for propaganda, or for the protection of special interests, by veto or otherwise, but not for the effective and timely formulation and support of large policy decisions. And without a strong legislature, special interest and ideological tendencies penetrate the bureaucracy, and undermine its neutral, instrumental character.

A fourth type of political system is exemplified by the Scandinavian and Low Countries. These systems appear to differ from the French and Italian in two respects. First, the party systems tend to be aggregative (e.g. the Scandinavian Socialist parties, the Belgian Socialist and Catholic parties). Second, the relations between parties and interests appear to be more consensual, which makes stable majority and opposition coalitions possible. Thus, though the party systems fail to aggregate interests as thoroughly as in the British case, the public policy-making function of the legislature is not undermined to the same extent as in the French and Italian cases. What appears to happen in the Scandinavian and the Low Countries is that the function of interest aggregation and general policy formulation occurs at both the party and parliamentary levels. The parties are partly aggregative of interests, but 'majority–minority' aggregation takes place finally in the coalition-making process in the legislature. This coalition-making process may be organised by parties in the formation of cabinets and the enactment of legislation or it may take the form of interest coalitions organised around issues of public policy. The capacity for stable majority–minority party coalitions and for relatively flexible issue-oriented interest coalitions is dependent upon the existence of a basic political consensus which affects both parties and interest groups. These appear to be the properties of the so-called 'working multi-party systems'.[6]

These simplified models of political systems represent some of the general theory of political groups shared by the participants in the planning session. It represents a set of hypotheses about the characteristics and consequences of the main types of interest-group systems to be found in the countries in which field research is planned. The problem of research tactics, of how to test, elaborate and develop our theories further remains to be specified.

The discussion stressed that the relation between political parties and interest groups could be viewed as a continuum with substantial autonomy at one limit and sub- and superordination at the other. The relationship patterns which exist in historical political systems always involve two-way flows of influence, which differ from one another in the dominant direction of the flow and the different patterns which are occasioned by different kinds of issues. Thus the extreme case of the Communist Party–Communist trade union dominance still involves a flow of information and influence from trade union to party, but the dominant direction of the flow is from party to trade union. In the case of the church and Catholic parties the flow of influence varies from country to country, and even among regions within countries. In addition, in certain areas of policy Catholic parties may be relatively free of church influence, or may even influence the Church to take a position consistent with or supportive of that of the party. In Germany, for example, the fact that the C.D.U. has both Protestant and Catholic support seriously limits the power of the Catholic Church to intervene in party policy-making. In other legislative fields, the freedom of Catholic parties may be sharply circumscribed by a rigid church position, as in the field of educational subsidies and the like. In other policy areas, e.g. social-economic, there may be more give and take in the relations between church and Catholic party.

These considerations suggested that analysis of the flow of influence between parties and interest groups would require not only an examination of the interconnections through financing, interlocking memberships and directorates, sharing of ideological beliefs and the like, but would also require a judicious use of case-study methods to discover the way in which different kinds of legislative issues affected the flow and pattern of influence between interest groups and parties.

It was generally agreed that a first task in analysing parties and

interest-group relationships would involve a study of the history of these organisations and of their interrelations. In some cases parties formed or were influential in the forming of interest groups (e.g. the Communists and the 'Red Unions', the Socialists and the *Force Ouvrière*); in many others interest groups formed or influenced the formation of parties (e.g. some labour parties by trade unions, Catholic parties by the church and/or elements of Catholic Action and agrarian parties by agrarian interest groups). Whatever the history of the relationship may be, it will provide an important body of evidence on the relationship between the two and changes in the relationship through time.

A second major research task would involve a careful analysis of the structure, internal politics, activities and ideologies of specific interest groups and parties as they relate to the flow of influence. (1) What is the relative financial strength of interest group and party? To what extent does the interest group finance the party, or vice versa? What is the special pattern of election financing? (2) What is their relative manpower potential? e.g. does the party depend on the interest group for personnel during elections as is the case for Catholic and labour parties, or does it have a manpower reserve of its own? (3) What is the relative effectiveness of the grass-roots organisations of parties and interest groups? e.g. is the party organised primarily at the centre and dependent on interest groups for the penetration of local areas, or does the party have an autonomous local organisation of its own? (4) To what extent is there overlapping in membership and officers as between interest group and party? (5) What are the factional and ideological characteristics of interest groups and parties, and how do they affect the pattern of influence?

VI. INTEREST GROUPS AND THE LEGISLATIVE PROCESS

Interest groups tend to seek out the important points of access in the legislative process, the points where legislative policy is initiated, and where revision, vetoing and favourable action are possible. Hence, the constitutional separation and distribution of powers, legislative organisation and procedure, the characteristics of the electoral system and the parliamentary party organisation, set the problem of interest-group access in the legislative process. Thus, the American system of federalism, and separation of powers, creates a different interest-group 'target structure' than does the British Parliamentary-cabinet system. The American federal system produces a party structure with its centre of gravity at the state level. This kind of decentralised party organisation limits the possibilities of congressional party discipline and hence opens the legislative process to interest-group penetration. The susceptibility of the legislature to interest-group penetration is enhanced by the American single-member-district electoral system which frequently exposes the legislator to the effective pressure of interests which may be especially concentrated in his constituency. In addition, the American separation of powers system grants a powerful and independent role in legislation to both the House and Senate. And since relatively large collegial bodies are, other things being equal, less able to aggregate interests and protect themselves against interest penetration than Cabinet-dominated parliaments, this aspect of American constitutional structure contributes significantly to interest-group action in the legislative process. If we consider this constitutional and statutory structure on the one hand, and the economic, regional, ethnic and religious composition of the American population on the other, it is hardly surprising that the penetration of the legislative process by interest groups in the United States is greater than in the United Kingdom. There, a unitary constitution and a Cabinet-dominated parliament make possible a disciplined parliamentary party system which protects the legislative process from effective interest-group penetration. The main targets of interest groups are the upper levels of the parliamentary and extra-parliamentary party structure where power is concentrated, and the bureaucracy. And because of the cohesion of the party system and the concentration of legislative power in the Cabinet, the impact of any single interest group – with the exception of the trade unions – is quite limited.

France presents yet another problem of interest-group access. In the United States aggregation and synthesis of interests is performed by the party system and a powerful presidency responsible to a national constitutency. France has neither a powerful executive nor an aggregative party system. A culturally and politically fragmented society choosing its legislators by means of proportional representation produces a legislature capable of producing only weak and unstable coalitions. The standing committees of the Assemblée are in many cases colonised by powerful

interests.[7] The net effect of this situation is a legislative process which can only rarely enact significant 'national-interest' legislation, but which regularly and characteristically protects and subsidises special interests. Still a fourth type of legislative interest-group pattern is to be found in the Scandinavian countries, where a stronger executive and a more aggregative party system limit the impact of interest groups in the legislative process.

These characteristics of constitutional, legislative and party structures affect not only the tactics of interest groups, but the very goals and objectives which they can reasonably attain. A disciplined party system and a powerful executive forces interest groups to direct their energies to the upper levels of the executive and the bureaucracy where only moderate claims, well supported with technical information, become possible. A non-aggregative and undisciplined party system as in France opens up the legislative process to covert interest-group domination of legislative committees and agencies, or to propagandistic interest-group manœuvres of which 'Poujadism' is only an extreme instance.

These hypotheses about patterns of interest-group–legislative relations in the European area suggest the importance of a careful analysis of the functioning of the constitutional, legislative, electoral and party systems as they relate to interest-group access. In other words, the aim of research in interest-group–legislative relations will be to determine the extent to which the parliamentary parties, or extra-parliamentary legislative institutions such as the American Presidency, are able to maintain independence of interest groups and relative freedom to legislate or influence legislation (*a*) by combining several interest groups in their support, and (*b*) by establishing and maintaining the discipline of the parliamentary party as a means of withstanding interest-group pressures.

The discussions in the planning session turned on the question of how to test these hypotheses economically and effectively. It was agreed that research here would concentrate on four problems.

1. The first of these is the analysis of the 'target structure' itself. This would involve an examination of the distribution and location of legislative power as between the parliament, the executive and the bureaucracy; the internal organisation and procedure of the parliament; the organisation and operations of the party system in the parliament; and the effect of the electoral system on the party system and on the position of the individual parliamentarian. This phase of research would also include an examination of the various devices for formal interest-group representation such as 'economic councils' and advisory committees and groups of one kind or another.

2. Closely related to this first task is a study of how interest groups perceive and appraise the 'target structure' of the legislative process. Such analysis would bring out significant differences between the interest groups in particular countries as well as between countries. If our general impressions are correct we would expect to find that British interest groups would share the same picture of legislative reality, and that it would tend to be a relatively accurate picture. In France, we would expect to find quite different conceptions of constitutional and legislative reality among interest groups. Thus, the different pictures of constitutional reality characteristic of the church and the Communist trade unions would be the consequence of basically different ideological presuppositions and political experience. The C.G.T. would view the French legislative process as incapable of realising working-class objectives. Consequently, their primary objective *vis-à-vis* the French Constitution would be to immobilise it, and to destroy it. The Church, on the other hand, would view the French legislative process as more amenable to pressure through access to the M.R.P. and right-wing parties, and through sympathetic elements in the bureaucracy. In many non-Western countries we would expect to find relatively new and inexperienced interest groups with highly inaccurate or vague pictures of political reality.

3. A third task involves study of the objectives of interest groups in the legislative process. Thus, in the United Kingdom we would expect to find interest groups entertaining moderate, secular objectives in the legislative process; in France we would expect to find interest groups seeking extreme, incompatible and rigid objectives; in non-Western countries we would expect to find many interest groups seeking 'expressive' rather than 'instrumental' objectives.

4. A fourth area of study is of the means or techniques employed by interest groups to attain their objectives – what we might call the interest-group style. Here, we would be concerned with whether or not the techniques are overt or covert, whether propaganda, negotiation, bribery or demonstration and riot are the

preferred means. In the United Kingdom we would expect to find an overt style, with propaganda, negotiation and bargaining the favoured means of interest groups. In France, we should find that much interest-group activity is covert, and that demonstration and bribery occur with some frequency. In many non-Western countries the style of interest groups may be largely covert, and the techniques heavily on the riot, demonstration and bribery side.

VII. INTEREST GROUPS AND THE BUREAUCRACY

The ideal of a neutral, instrumental bureaucracy, responsible to the political arm, comes closest to being realised in political systems where the political culture tends to be secular and rational, and where the functions of political choice – that is, the articulation and aggregation of interests and the making of public policy – are performed by specialised agencies. If the interest-group system fails to articulate and communicate the demands of significant sectors of the population, or if the political demands which it transmits are irreconcilable, then the capacity of the party system to aggregate and to compromise demands and facilitate political choice in the legislative process is undermined. If the parliament and the 'political executive' fail to perform the functions of interest articulation and aggregation, then the problem gravitates to the bureaucracy, and its capacity for neutral, responsible administration is weakened. Under these circumstances the bureaucracy tends to be multi-functional. It articulates and aggregates interests, makes public policy and administers it. To be sure, the so-called 'neutral bureaucracies' in the West do these same things, but only to a limited extent and under control, and in situations where there are effective political agencies outside the bureaucracy.

For purposes of the planning session four types of interest-group–bureaucracy relationships were discussed. The first of these is illustrated by the situation in the United Kingdom where moderate interest groups, representative of the significant interest tendencies in the population, are aggregated by disciplined parties. Both of these conditions make possible an effective organisation of the legislative process, and an effective political control over the bureaucracy. The second is illustrated by the Scandinavian countries where aggregation of interest is less thoroughgoing than in the United Kingdom. In the Scandinavian countries aggregation occurs both in the party system and in the coalition-making process in the parliaments. Nevertheless an effective parliamentary process makes possible a neutral, instrumental bureaucracy. The third type is illustrated by the situation in France where effective aggregation of interests and policy-making does not occur in the party system, the parliament or the bureaucracy. The net result is a political system in which the agencies of political choice fail to function, and in which basic policy decisions cannot be made. It tends to be a government of 'protection and maintenance' in which the effective agencies are the interest groups and the bureaucracy. The bureaucracy cannot be a neutral, responsible instrument under these circumstances. It is 'colonised' by interest groups and penetrated by incompatible ideological tendencies.

The non-Western areas present a variety of interest-group–bureaucracy relationships. In one quite common pattern parties and associational interest groups are only in their beginnings, and the political process operates primarily within the bureaucracy itself. The bureaucracy is a congeries of interests – family, religious, status, class, professional, ideological and the like. A bureaucratic or army clique makes informal coalitions of these interests and maintains contact through them with their counterparts outside the bureaucracy. The political process may take the form of an unstable pluralism, or in a second variety a particular clique or individual may establish authoritarian control and operate with the backing of the army. Another non-Western pattern is one in which an authoritarian party is formed, one of its main functions being to mobilise groups in the society, articulate their interests and bring them to the support of the ruling groups. If the tendency of the authoritarian leadership is conservative, then traditional interests – status, economic, religious – will have favoured access to the bureaucracy. If the tendency is towards modernisation, then urban, professional, industrial and labour interests may have favoured access to the bureaucracy, and the traditional interests may be in a disadvantaged position.

The planning sessions agreed on the following specifications for research on interest-group–bureaucracy relationships: (1) Each study would have to appraise the substantive areas of economic and social life affected by the bureaucracy. This would indicate the kinds of interests which were being affected by bureaucratic activity, and the ways in which they are being affected. (2) Each

study would have to examine the organisational and influence pattern of the bureaucracy. To what extent and by what means is the bureaucracy controlled by extra-bureaucratic agencies such as the cabinet and/or the parliament? To what extent is the bureaucracy internally cohesive and subject to central control? Or put the other way around, to what extent are specific agencies and sub-agencies independent? Needless to say both the external political and internal control patterns would set some of the problems of access of interest groups to the bureaucracy. (3) What are the characteristics of the members of the bureaucracy, particularly its higher echelons? From what social groups are they recruited? What are the training patterns, and what kinds of skills do they possess? What are their ideological and partisan characteristics? All of these attributes would affect their susceptibility to interest-group pressures. (4) What are the goals and tactics of the various interest groups *vis-à-vis* the bureaucracy? Are they primarily concerned with gaining specific family, status group, class or professional advantages, or do they advocate diffuse ideological objectives, or both? What methods of influence do interest groups employ in relation to the bureaucracy? Is bribery or violence a common pattern? Or is propaganda, negotiation and the provision of information about the consequences of bureaucratic action for specific interests the typical pattern? (5) To what extent do interest groups actually penetrate the bureaucracy? Here we are not only concerned with formal provision for interest-group representation in specific agencies in advisory capacities, or with the special relations between administrative agencies and their particular constituencies (e.g. labour ministries, economic ministries, agriculture ministries and the like), but also with the direct control over specific bureaucratic functions, in other words the 'colonisation' of the bureaucracy by interest groups.

VIII. COMPARISON OF WESTERN AND NON-WESTERN INTEREST GROUPS

One of the most challenging questions confronting the planning sessions was the problem of comparing Western and non-Western political systems. Are they so different in culture and practice as to constitute basically different political species? Was the approach of political science scholarship in the study of Western and non-Western politics so different as to rule out effective collaboration? Did the group participating in the discussions and planning field studies share a common theory of politics which would make collaboration possible? The discussions led to the conclusion that Western and non-Western political systems have much in common, and that the particular group of scholars present shared a common approach and theory of politics.

The basis of this common outlook might be called the 'functional' approach. The participants in the comparative study agreed that they were concerned with the functions of political choice, and with the ways in which these functions were performed in different societies. Every independent society makes political choices, i.e. broad policy decisions which are backed up by severe sanctions. In making and enforcing these political decisions all societies have some way of articulating and communicating political demands, aggregating these demands, translating them into choices of political personnel and public policy, executing these decisions in specific cases and testing the appropriateness of these specific actions. In studying interest groups comparatively the participants in the study are primarily concerned with the structures, institutions and processes by means of which these functions are accomplished. Research conclusions as to which structures and processes perform which functions in different societies, and how they perform them, will provide the basic materials for comparative analysis.

Not only are these functions performed in all independent political systems – Western and non-Western – but the structures and processes which perform them in both areas overlap to a considerable extent. The West is more like the non-West than we sometimes think. Even in the most differentiated and specialised political systems in the West, such interest groups as families, status groups and religious communities affect the political process. And in most of the non-Western countries – however 'underdeveloped' – the beginnings of functionally specialised political parties, and associational interest groups such as trade unions and trade associations, may be found. Even in the field of political communication the highly elaborated mass communication systems of Western societies should not obscure the fact that informal and face-to-face communication is still a political factor of enormous importance.

In still another respect, Western and non-Western systems are alike. While it is true in general that Western political structures are more specialised than the non-Western, there is much 'multi-functionalism' in the West. Thus in a country such as France political parties and interest groups are not sharply differentiated from one another. And in all countries the structural specialisation of policy-making and administration is by no means complete, nor can it ever be complete. If it is peculiar to non-Western countries that bureaucracies are penetrated by interest groups and ideological tendencies, this situation differs only in degree from the Western pattern where rationality, responsibility and neutrality are only partially realised at best.

The comparative study of interest groups will include a number of cases in Asia and Latin America. In particular it is anticipated that there will be studies of Malaya, Burma, India and Indonesia in Asia; and of Brazil, Mexico, Cuba and Peru in Latin America. The non-Western specialists at the planning session agreed to follow much the same pattern as the European specialists. This would include a general appraisal of the kinds of interest groups to be found in these countries, their composition, size, goals, methods of influence and the like. It would also include more detailed investigations of selected interest groups. These intensive studies are likely to include trade unions, peasant groups, student groups, religious and tribal or ethnic groups. Some attention would be paid to the 'mob' as an 'anomic' interest-group phenomenon, resulting from lack of effective access of some social groupings to politics, and generally poor political communications.

The planning session at Palo Alto was followed by another held in June 1957 at Michigan State University for a number of scholars about to undertake research on political groups in Latin America. A third is planned for the late spring in 1958 for recipients of grants for field research on political groups during the present year. By setting common questions and discussing common approaches, the Committee on Comparative Politics hopes to introduce greater coherence in the research efforts under way. More recently it has begun to plan a series of comparative analyses which will draw the findings of these and other research efforts together.

Notes

1. Maurice Duverger, *Political Parties* (New York, 1954) pp. 392–3.
2. David B. Truman, *The Governmental Process* (New York, 1951) pp. 14 ff.
3. Lucian W. Pye, *Guerrilla Communism in Malaya* (Princeton, N.J., 1956) pp. 346 ff.
4. The report on this topic was presented by Sigmund Neumann of Wesleyan University and Val Lorwin of the University of Oregon.
5. Kabin, Pauler, Pye 'Comparative Politics of Non-Western Countries', *American Political Science Review* (Dec. 1955), pp. 1022 ff.; Almond, Cole, Macridis, 'A Suggested Research Strategy in Western European Government and Politics', ibid., pp. 1042 ff.; Almond, 'Comparative Political Systems', *Journal of Politics* (August 1956), pp. 301 ff.; Francis X. Sutton, 'Social Theory and Comparative Politics' (unpublished paper); Samuel H. Beer, 'Pressure Groups and Parties in Britain', this *Review* (March 1956), pp. 1 ff.; Dankwart Rustow, *Politics and Westernisation in the Near East*, Center of International Studies, Princeton University, March 1956; Rustow, 'New Horizons for Comparative Politics', *World Politics* (July 1957); Sigmund Neumann, 'Comparative Politics: A Half Century Appraisal', *Journal of Politics* (August 1957); Gabriel Almond and Myron Weiner, 'A Comparative Approach to the Study of Political Groups', Agenda Paper for the Dobbs Ferry Seminar of the Committee on Comparative Politics, June 1956.
6. See Dankwart Rustow, 'Scandinavia: Working Multi-party Systems', in Sigmund Neumann (ed.), *Modern Political Parties* (Chicago, 1956) pp. 169 ff.
7. Henry W. Ehrmann, *Organised Business in France* (Princeton, Princeton U.P., 1957) ch. v.

In previous chapters the State has been taken for granted as the necessary setting for government and politics. In this section we look at the way government and politics take place without the State.

There are a number of non-industrial societies, both in the past and the present, which have functioned without the institution of the State. Though the detailed way in which this works varies from society to society, the basic principle is the same: the functions of politics and government are, in one sense at least, performed by other institutions than the formal machinery of the State.

The Tiv of Northern Nigeria are one example of a stateless society. They not only have no state machinery; in addition they have no formal leaders or officials, so that their society is marked by 'the absence of offices and office-holders'. But nevertheless they do have a kind of politics and government. Through the political units knows as 'lineages' the function of a political system is fulfilled, in Bohannan's words, 'to control law and warfare – the exercise of power – within a territorial framework'.

This is paralleled in industrial societies with fully developed state political institutions. Social control and decision-making can take place without involving the formal machinery of the central state or statelike machinery at local level. In the American shopping centre the group functions like a 'little community' on its own, in which unwritten laws and personal relationships are used to resolve conflict and preserve social relations in a way akin to the 'law of primitive communities'. In the queue, a miniature political system can be detected which regulates behaviour and controls conflict.

SOCIAL AND POLITICAL ORGANISATION OF THE TIV

Paul Bohannan

SOCIAL ORGANISATION

The lineage system

The social organisation of the Tiv is simple. It is, in fact, so simple that it is difficult to understand. It utilises a single principle of organisation – the agnatic[1] lineage structure, based on the principle of segmentary opposition. The meaning of these terms will become clear as the exposition proceeds. The family has at its core the agnatic lineage, the compound also has at its core the agnatic lineage, and the lineage is fundamental to the political and religious institutions.

A lineage is a group of people who are descended from a common ancestor or ancestress. An agnatic[1] or patrilineal lineage is a group of people descended from a common ancestor through males only. It is a ubiquitous group among the Tiv. The matri-

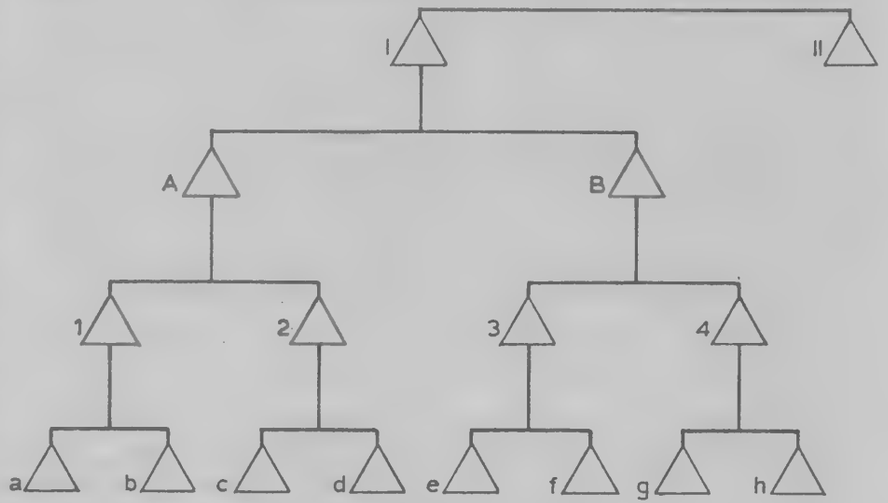


Fig. 26.1 Genealogy of Tiv lineages

lineal lineage is not recognised by the Tiv for any purposes whatever.

A lineage *system* comes into being when several lineages are grouped to form a single social unit. There are several principles on which such an organised agglomeration of lineages may be formed. The one used by the Tiv is the principle of segmental opposition. They, like some other African peoples, organise their lineages into such a system by citing genealogical associations. The lineages made up of the descendants of men who were brothers

form an inclusive lineage in the name of their common 'father', as opposed to lineages descended from more distant kinsmen. Thus, as Fig. 26.1 shows, all Tiv can be placed on a single agnatic genealogical chart and at every generation back in time a new and inclusive lineage is formed. Ultimately, all Tiv form a single lineage, descended from the original ancestor, 'Tiv'.

The principle of segmental opposition is expressed in Tiv terms by saying, 'Would I take something away from my full brother when I can get it from my half-brother?' In scientific terms we explain it by saying that two segments of any lineage, whether those segments be individual people or smaller lineages within larger lineages, are completely equal in all things. Two lineages opposed in one context may join forces in another as a single lineage against an external lineage or against foreigners.

To take an example, Fig. 26.1 represents a genealogy in which only the males have been shown. Each symbol can represent either a living individual or a lineage of agnatically related individuals. Thus, if *c* were to argue with *d*, the two, be they lineages or individuals, are equivalent to one another; no one else is concerned in their dispute. However, if *c* were to have his dispute with *b*, then *c* and *b* are themselves not equivalent. Therefore, *d* is obliged to come to the assistance of *c*, and *a* is obliged to come to the assistance of *b*. In effect, the dispute then becomes one between the descendants of 1 and the descendants of 2. At that level, 1 and 2 are equivalent, and no one else is concerned. In the same way, if *c* were to be in dispute with *f*, *c* is part of Lineage 2, which is itself part of Lineage *A*, whereas *f* is a member of *B*. Therefore the dispute would be carried to the level of the opposed equivalents: Lineage *A* and Lineage *B*. In an ordinary situation, all the members of the two lineages would be opposed to one another over any dispute between their members. It is obvious that this principle can be carried all the way to the top of the genealogy, no matter how great its genealogical depth: the members of Lineage I are opposed to those of Lineage II, and so on.

The genealogies that support the Tiv lineage system are about seventeen generations in depth, and go back to a single ancestor. Every member of the tribal group can trace his descent, real or putative, from the founder of the entire group. He does so by tracing the relationship through males only.

This means of linking lineages into systems is remarkable in that it is such a simple mode of organising the system at the same time that it disallows a highly developed institutionalised authority. If one person is given more authority than another, the lineage system ultimately breaks down and will not work: its segments are no longer equivalent. If authority is vested in a man from Lineage *A*, all others can ultimately dissociate themselves by invoking the fissions built into the system.

Such a lineage system forms the basis of the family and the household system of Tivland; it also forms the basis for the settlement pattern and the political system. We shall take these matters up in that order. Such a lineage system must occasionally be subverted, and we shall examine some of the recognised means for doing so.

The lineage among the Tiv forms the basis for all local groups. Tiv are largely patrilocal – a man lives with his father. In three samples taken from widely different parts of Tivland, 83 per cent of males were living with their lineages. Some, however, did live with their mother's agnatic lineages, a few with their mother's mother's or father's mother's agnatic lineages, and 2 per cent were in areas where they were considered 'strangers' – areas to whose lineages they could trace no kinship at all. But residence is uncomprisingly virilocal. I discovered only three men who were living uxorilocally, and of these one was in the record of a murder trial of a man who had killed the person who taunted him for living uxorilocally.

The exogamic group is a lineage containing four or five hundred people. It is sometimes possible for Tiv to take lovers within this group so long as they do not contravene their society's definition of incest: anyone who shares a single grandparent is forbidden. There is another lineage, more inclusive than the first, beyond which it is possible to take a senior wife or one who can ritually be more fully honoured (though she is no more legitimate). . . .

POLITICAL ORGANISATION

The lineage system is the basis not only of the family system but also of the geographical and political organisation. The geographical organisation is a simple one. At one level in the lineage structure there is associated with each lineage a territory within

which, as we noted, 83 per cent of the males are agnatic members of the lineage.

As Fig. 26.2 shows, if the territory of Lineage *a* is beside that of *b*, in combination they form the territory of Lineage 1. In the same way, if Lineage 1 has its territory beside that of Lineage 2, made up in its turn of Lineages *c* and *d*, then Lineages 1 and 2 form the lineage, and the lineage area, of Lineage and area *A*, and so on. Therefore, the geography of the country is seen in accordance with the same pattern or model as is used for the social organisation of the people themselves in other contexts. Tiv have place names for streams and for the highest hills. Aside from this, all place names are referred to in terms of the lineages of the people who occupy them.

The individual's geography becomes an integral part of his kinship system. One's agnatic lineage, of whatever depth, is called his *ityô*. My mother's *ityô* is my *igba*. My mother's mothers' people are my *igba ngô*, and my father's mother's *ityô* are my *igba ter*. During the early stages of my life, my geographical world is limited by my various *igba*, and by my *ityô*. Every lineage to which I belong right up the scale of lineages is my *ityô* until I come to a point at which my *igba* and my *ityô* are structurally opposed segments. Therefore, if I am a member of Lineage *b*, and Mother has her *ityô* in Lineage 4, then my *ityô* is Lineage *A* and my *igba* is Lineage *B*. However, if my mother comes from Lineage 2, then that is my *igba* and Lineage I is my *ityô*. In this way, the entire world with which I am acquainted as a youngster can be turned into a series of kinship areas.

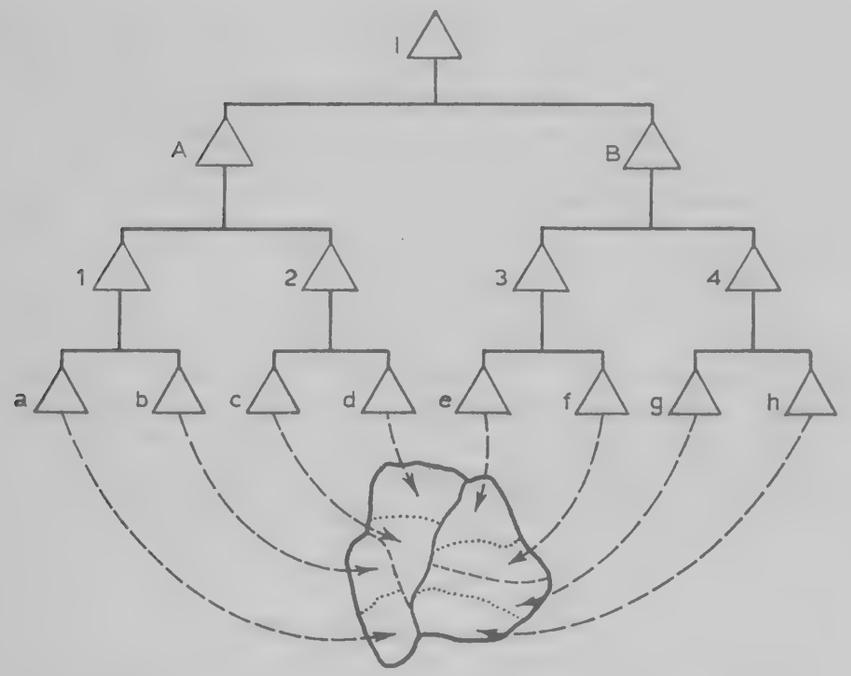


Fig. 26.2 Reprinted from P. Bohannan, 'Map and genealogy of Tiv political divisions' (*Africa* 24, 1954) by permission of the International African Institute.

The function of a political system is to control law and warfare – the exercise of power – within a territorial framework. Since the territorial framework is conceived in lineage terms, the control of power must also take these terms into consideration.

The nature of the lineage system based on the principle of segmental opposition is probably nowhere better shown than in the way in which it deals with the problem of warfare. It acts as a system of built-in checks and balances that generally allows warfare to take place without getting out of hand. In theory – and usually in fact – fighting is stopped short when it embroils two equal and opposed lineage segments. When Lineage *b* fights with Lineage *c* all of Lineage 1 and Lineage 2 may become involved, but Lineage *B* will not because the members are not more closely related to one of the combatants than they are to the other. Individual people of Lineage *B* may, however, go to their matrilateral kinsmen in either Lineage 1 or Lineage 2 and join, as kinsmen, with the combatants. Such a situation does not count as the official entry of Lineage *B* into the fray, and, indeed, if individuals from Lineage *B* do join, there are likely to be as many on one side as there are on the other. Thus, by means of the lineage principle, wars are localised.

Close kinsmen and even lineage mates may be fighting on opposite sides in any given fray. When asked what they do if brought face to face with their lineage brothers in the fighting, Tiv reply that they see one another and shout, 'Hey, look, it's me. I'm your brother. You go that way, and I'll go this way.' So that although they are fighting on different sides, they do not engage one another in hand-to-hand combat.

Law

Like warfare, law was also institutionalised within the limitations imposed by the lineage system. The two most important characteristics of the legal practices are both imposed by the lineage system: the absence of offices and office-holders, and the fundamental equality of opposed lineages.

In the indigenous political system, there is no role above the level of compound head that can be called by a title to which authority in any sense may inhere. Political leaders can be described, and the functions of leadership discussed – but any man who performs the tasks of leadership can be described with the words. The words describe acts, not roles. During the colonial era, one of the difficulties was a constant misunderstanding on the part of the British administrative officials, who kept assuming that these descriptive words were in fact names for official positions. The British then tried to tie down each of these 'positions' to a given individual, thereby creating a static system of the sort to which they were accustomed. Tiv refused to have power concentrated in the hands of any given individuals, with or without offices. The few exceptions to these generalisations were ritual offices that Tiv bought from the Jukun tribe nearby, and in which little or no political authority inhered.

In the absence of officials, law was maintained by a series of meetings of the elders of lineages of various depths. The lineage involved in the settlement of a dispute was determined in precisely the same way as was the lineage involved in fighting. The dispute settlement also became a general airing of grievances between the two contending groups, not merely the particular difficulty that was instrumental in bringing the elders of the two groups together. These sessions, which can be called moots, were a device by means of which all the difficulties involved in the community could eventually be brought into the open so that, at least, a general awareness of them could be had and, at best, practical solutions found. All of the elders of the community participated in these moots. There was no specialised personnel such as there is in a court.

The British established a system of courts among the Tiv soon after World War I. They took lineages that contained about 8000 people and gave them, quite arbitrarily, the term *clan* – this term is not used in accordance with any anthropological definition. Each clan, they said, was to have a 'chief', who later came to be called a *ter* (father). He was to be assisted by three or four judges; again the British administration took a Tiv word, *mbatarev*, which described men who were influential in settling disputes within their community, and made of it a title for the officials concerned. Each of the *mbatarev* (chief and judges) was given two or three Native Authority policemen, armed only with nightsticks. There was also a scribe attached to the chief, both to write letters for him and to keep rudimentary court records. Such an organisation was called a Grade D court.

There was another court, made up of from four to as many as a dozen combined Grade D courts. They were called Grade C courts, and acted on appeals from the Grade D courts. They also acted as courts of first instance in civil cases involving over £25 and criminal actions which bore prison sentences of between three and six months. The combination of all the Grade C courts of Tiv Division (this obviously is also the collection of all the Grade D courts) was called the *jir tamen* (big court). It acted as a Grade B court and was the appeal from the Grade C court. There was no Grade A court among the Tiv – there were only a few in all of northern Nigeria. Appeal from the Grade B court was to the magistrate's courts in the system of law imposed by the British on the Northern Region of Nigeria.

The judges of a Tiv court all sit together to form the bench. The complainant is allowed to recite his complaint – he may go on at any length so long as he can convince the judges that what he is saying is relevant. The defendant is then allowed to tell his side of the story. Neither of the principals is sworn, or allowed to take an oath. Witnesses, however, may be sworn, but often are not. They are called by the judges, and although a principal can suggest witnesses, he cannot call them. After the stories of principals and the evidence of witnesses are all in, the judges announce a decision. They do not speak in any particular order, and they may disagree with one another, though disagreement will be overcome before a final decision is announced. The decision must meet the concurrence of the principals and of the community – community pressure can make a recalcitrant litigant accept a decision, and there is no other sanction that is fully effective. Decisions are *ad hoc*, with minimal reference to legal precedent or to any body of laws, although the ethnographer can elicit both precedents and laws.

The major work of any Tiv tribunal is made up of the settlement of marriage disputes, the filiation of children and disputes over land and farming rights. Tiv are a litigious people and enjoy court cases; they also get a great deal of satisfaction from having their difficulties arbitrated in this way. The judges, who have considerable leeway to make compromises, give moral lectures and ethical advice. Although they are not always orderly, Tiv are fundamentally a law-abiding people.

It can be seen that the administrative courts follow the hierarchy of the Tiv lineage system. At the same time, however, the courts allow an individual some degree of escape from the dominating principle of the lineage. Two other institutions provide an even better means of escape if the lineage principle becomes too confining. One of these is the age set; the other is the market-place.

Age sets

Tiv are organised into age sets called *kwagh*. Each set includes men born in a period of about three years, but full brothers – even twins – cannot belong to the same age set. Like most other Tiv institutions, the age set does not originate formally, but rather drifts into existence as it is required by members of the group who will be its adherents. By the time that men reach the age of seventeen or eighteen, they know who are going to be in their age set; the group of young men is taken in hand by men from an older age set – but younger than that of their fathers – who are called *igba* to the younger age set. The *igba* instructs the members in their duties to one another. These duties involve mutual assistance in work and in financial matters, but, more important, they involve the capacity to treat with diviners, doctors and other ritual specialists. The age set is a man's primary ally should the members of his lineage withdraw supernatural protection, which is evident to Tiv whenever a person falls seriously ill. The age set, which includes all of one's age mates among one's lineage of whatever depth, can convene all of the lineage members who belong to older age sets to inquire into the ill health or discomfiture of one of its members. Therefore, if the members of the lineage become too pressing and do not allow a person full rights or if they threaten him, it is to his age set that he turns. The age set is thus one of the chief alleviants of lineage pressure.

Age sets are of greatest importance to young men in their early twenties, who are doing a major part of the physical work of farming. Age mates very often work together on the farms first of one, and then another of their members. The recipient of the labour for the day provides food and beer and is, of course, obliged himself to work for another member of the age set on another day.

The primary function of the age set changes when men reach the age of about forty. At that time, when the members are passing into elderhood, the age set is used most to maintain rights and to stave off the pressures of other members of the lineage who may be jealous, or who may resist the advancement or the wealth of the incoming group of elders. After its members reach the age of about fifty, their age set becomes very little more than a sentimental association.

Other counter-lineage factors

It is readily seen that Tiv social organisation gives little opportunity for men to achieve legitimate power positions. There are several avenues, however, by which some power and prestige may be gained. One of these is to become the most important man in the age set; another is to become an important ritual specialist; and still another is to gain control of a market. In order to maintain one's position in a market-place, it is necessary to be able to assure those who come to trade that they will be safe and that their rights and goods will be respected. In order to make such assurances, it is necessary to have fairly large and fairly strong followings. These two factors interplay, and if a man is canny enough to do so, he can become a power in the land by controlling a point of trade and communication. He can, in this way, even stand out against his own lineage. Hence the market becomes one of the most important of the alleviating factors for the force of the lineage principle.

The absence of offices is paralleled by the fact that Tiv is a classless society. Although traditionally there was a group of slaves, the slaves themselves were kinless persons attached into a lineage and into a family by a slave–master relationship instead of by a kinship relationship. Within the family, they suffered an inferior status, but before members of other families, or before members of other lineages, they were considered as equal to lineage members by birth. Slaves, if they were the oldest men in the compound, could

inherit the compound headship, although such an arrangement usually was not stable and led to a breaking-up of the compound at a relatively early stage. Slavery in Tivland was never, even secondarily, an economic institution, and was never widespread.

The only hierarchical social relationships are those that involve kinsmen and most specifically those that involve older and younger persons. Older men are always right by virtue of their seniority, and when strange Tiv meet, they must soon decide which is senior.

Friendship between unrelated men may be institutionalised and they may or may not become blood brothers. These special friends have specific obligations and rights, some of which can be enforced in a court or a moot. They include financial assistance and assistance in labour; perhaps most important of all, the best friend is a sanctuary if one must escape the machinations of one's lineage. It is a fact that friendship in Western society is an extremely personalised relationship with a minimum of formal content. For this reason, Western social scientists have given little regard to the institutions of friendship in other societies. Although best friends among the Tiv like each other and do personalise their relationship, the formal element of mutual assistance and protection is considered to be the more important aspect.

We have seen that Tiv political organisation is oriented about the lineage principle. It is inevitable that whenever a single principle of social organisation emerges so strongly, difficulties emerge. As we observed above, Tiv have several means of getting around such problems: the market itself is one such organisation that can be used to control the inexorability of the lineage principle. The age set is another. Various kinds of co-operating groups or 'friendly societies' and several types of institutionalised friendship also ameliorate the force of the lineage principle, thus, in fact, making it possible for the principle to work.

Note

1. Recruited or related on the father's side; patrilineal (Ed.).

DISPUTE SETTLEMENT IN AN AMERICAN SHOPPING CENTRE: A PRELIMINARY VIEW

Spencer MacCallum

An impressive development in modern American land tenure has been the post-war rise of professional property management. Besides this, there has been a proliferation of new forms and functions of the multiple-tenant income property. Types of the latter that have become familiar in the landscape since the war include shopping centres, marinas, industrial and research parks, medical clinics, professional and office centres, mobile-home parks, real-estate complexes, for which Rockefeller Center was an early prototype, and some of the new 'planned communities'.

As a landlord–tenant arrangement, the multiple-tenant income property represents a land-tenure pattern of a kind we are used to thinking of in terms of peasants and feudal institutions and of absentee aristocracy battening on the land. But in the present case, the landlords are modern, efficient firms specialising in property management for income on invested capital. From a structural viewpoint, nevertheless, these organisations are indeed akin to the landed estates and manorial organisations of antiquity, in which the internal public authority derived mainly from the proprietary land authority. They are truly 'little communities' in modern garb, for to some extent they are modern counterparts also of village communities organised on the kinship tie, in which the land-distributive function was an attribute of the public authority, such as there was, vested in the chief or headman of the group.

The point I wish to make in this paper is simply that the multiple-tenant income property can be viewed in its internal organisation as a community of landlord and tenants. Of the many forms, the shopping centre is sociologically most interesting because of the complex relations among its members. This complexity derives especially, but not by any means solely, from the need for joint promotions on which the centre characteristically depends because of its locations away from downtown where there is sufficient 'natural traffic' to sustain the stores. While neighbourhood centres are narrowly specialised, the larger shopping centres – containing from 50 to 200 merchant tenants and a mixed composition of land uses, including professional and office build-

ings and even motels and apartments in the same plan of development – begin to resemble communities as we are accustomed to thinking of them.

Pospisil[1] has pointed out that we need not think of there being one system of law in a given society, but that each subgroup was its own.[2] Following this idea and inquiring into the law of the shopping centre as a community of landlord and merchant tenants, we find an intriguing situation: The legal system of a shopping centre is composed of two parts, a written part and an unwritten part. For the totality of the leases in effect at any given moment in time is the written, formal law which defines the respective rights and duties of all parties, merchants and landlord alike. The employed personnel of the individual business firms within the centre do not come directly under this jurisdiction, but are subject to the rules of their respective organisations. They belong to the shopping centre community not in their own right, but through their respective employers, as the members of the domestic groups making up a village community might be related to the polity through their respective 'patresfamilias' and not in their individual rights. Case 3 below illustrates an appeal to lease law, and Case 4 the hierarchy of legal levels in the centre.

The informal law on the other hand develops outside of the leases and consists of a body of rules and understandings, by-laws as it were, governing behaviour in the centre. Case 5 suggests one of the ways in which such laws may develop, as does Case 4.

An important problem of shopping-centre administration is writing effective leases. One of the requirements of an effective lease is that it consists with the existing leases – the rest of the body of written law – and also with the unwritten law of the centre. A general development in the shopping-centre lease has been a movement away from the traditional form of lease which defined a narrow dyadic relationship, specifying what the two parties would do for one another and leaving it at that, towards the lease becoming a conscious instrument of social policy in the shopping centre as a community of merchants. Increasingly, it has been developing lateral extensions, as it were, citing positive obligations of the lessee not only to the landlord but also to the other merchants in the centre. Clauses requiring tenants to participate in the merchants' council, spend a minimum percentage of gross income on advertising and a portion of this in council-sponsored media, co-ordinate his minimum opening hours with those of other merchants, enforce the centre parking rules among his own employees, and so forth, are becoming standard items in shopping-centre leases.

If the shopping centre is thought of in this way as an autonomous community of landlord and merchant tenants, then many suggestive parallels come to mind between its law and the law of primitive communities. The emphasis in both is towards resolving differences and preserving social relations, and neither is fundamentally concerned with rules and penalties.

One of the most fruitful areas for study, however, may be the problem of writing effective leases, since it is here that we confront the problem of the relation between written and unwritten law, and perhaps also the problem of boundaries between law and custom, on a manageable scale and, incidentally, without a language barrier. Moreover the recent growth of these new forms of organisation on the land – truly estate forms of urban land tenure – appearing *de novo* in the bare space of twenty years, affords unusual conditions for studying institutional emergence and change. For the anthropologist, moreover, it offers a natural opening to the study of contemporary society.

CASE 1: THE CASE THAT TURNED AROUND

Informants: Shopping-centre manager; former promotion director; succeeding promotion director; consultant from an outside promotion firm who attended meetings of the board of directors of the merchants' council after the resignation of the promotion director.

Facts: Each year the baker had made a 500-lb. cake for the centre's Birthday Anniversary Sale. The new promotion director thought the baker charged the merchants' council too much for his cake. He told the baker that and asked for a specific bid on the following year's cake. When the baker failed to give a price, the promotion director obtained a cake at a low price from a baker firm in another city that operated a chain outlet for ready-baked goods as a concession in the supermarket in the centre. At a council meeting after the Anniversary Sale, the baker was irate. He said the arrangement had been 'rigged' and that he had been unjustly accused of charging too much for his cake. He said his

cake was better quality. He charged that the promotion director had been disloyal to the merchants in the centre by going 'off the mall'.

Outcome: The promotion director publicly asked the baker for his friendship and was refused. The baker withdrew from the council and stopped paying dues. Within a year, the promotion director had resigned under diffuse pressure from the council. The baker still did not return. Shortly afterwards, the council directors decided to feature a cake again at the next Birthday Anniversary Sale and to ask the baker to provide it.

Comment: The former promotion director's account differed from the other accounts on the point of whether he had 'gone off the mall' or whether there were, as he recounted it, 'two bakers in the centre' (e.g. the centre manager's story: 'We went outside and got another baker . . .'). He thus contrived to avoid the issue of disloyalty on a technicality and cited more fundamental reasons for his inability to get along with the merchants, of which he said the baker's withdrawal and refusal to pay dues was only a symptom and a further aggravation. Among his reasons, he cited (1) tension in the centre over a levelling-off of volume gains over the past year due to new discount stores in the area and a natural tendency for a centre's growth to slow around its seventh year; (2) the structural problem of his being employed directly by the council instead of by the landlord, so that every merchant felt he 'owned a piece of' the promotion director; and (3) that he was not temperamentally suited to the job.

He predicted the problem with the baker would be settled by a delegation from the promotion committee of the council going to him and saying, 'Frank, we like your cake best and want to have it, to hell with the cost. Will you make it?' And Frank would say, 'Well, I'll think it over.' Two months later, just prior to the writer leaving the field, this seemed to be the way events were developing; the promotion committee had decided to ask the baker to bake the cake, without making a point of price.

The succeeding promotion director suggested a further element in the baker's stand may have been that it gave him a chance not to pay dues. He quoted the baker as once having said to him, 'Promotions don't help bakers.'

CASE 2: THE CASE OF THE SHOPPING CARTS

Informant: Shopping centre manager.

Facts: A supermarket's checkout area faced on to an arcade of shops. On more than one occasion, boys bringing in shopping carts from the parking lots made up lines of carts that partly blocked access to shops in the arcade. Several of the merchants spoke to the manager of the centre about it.

Outcome: The manager took the problem to the food-store manager and questioned him in detail about what could be done. Together, they worked out a different system of stacking the carts and decided upon various areas of the store where excess carts would be stored when not in use. The manager of the centre then spoke to the other merchants, telling them something had been worked out, and suggesting they be lenient with the food-store manager, pointing out that often it is not the manager but employees who make the difficulties.

Comment: The centre manager emphasised that problems are handled in a face-to-face manner when they are still small. He said the management could do this because they spent so much time with the tenants that they learned the characteristics of each person. He said, 'Just so they know you're doing something, it doesn't matter what you do. The merchants have got to feel that their interests are being looked out for.'

CASE 3: THE CASE OF THE WINDOW FULL OF SLIPPERS

Informants: Shopping-centre manager; promotion director.

Facts: A shoe store had an exclusive to sell shoes in the centre. The ready-to-wear shop therefore had an express clause in its lease forbidding sale of shoes. The shop bought a lot of colourful, mule-type slippers of many different kinds and filled a window full of them. The shoe-store man did not make a point of going to the manager of the centre, but he met him on the street, walking casually, and said, 'I see they're selling shoes next door. I didn't think they were supposed to do that.' The manager said, 'I'll chase the lease up and see what I can find out.' He talked to the manager of the ready-to-wear shop, who said, 'Oh no, these aren't shoes. They're boutique.' The centre manager went to his

office and 'let it cool off a bit . . . probably it was two hours'. In the meantime, he looked for *boutique* in the dictionary and did not find it. He went to the shoe store, then back once more to the ready-to-wear shop, talking the thing down, rather than inflaming it. The ready-to-wear man, appealing to the professionalism that is stressed in the shoe business, said it really was not a shoe because it did not have to be fitted. It was not customised. He said, moreover, that he had a tremendous stock.

Outcome: The centre manager went to his office, and then, after a while, went back to the manager of the ready-to-wear shop and 'just sat down with him'. He said, 'This is a borderline case; you're asking me to define something the dictionary hasn't defined. I think the thing to do on this is to compromise. Instead of filling the window and making a specialty item of it, just put two or three pairs in the window as you would an accessory. Then you won't look like a shoe store.' He allowed him to keep his stock and sell it.

Comment: The centre manager said, 'We don't let those things go quickly. We let them cool, go slowly on them.' Commenting further on how cases usually come to his attention, he said, 'The grapevine is faster than your ears . . . I'm one of the last people that ever gets a complaint. About 90 per cent of the complaints I get are indirect. Somebody says, "So-and-so's been beefing about that." '

CASE 4: THE CASE OF THE OBSTINATE NURSE

Informant: Promotion director.

Facts: Despite warnings placed on the car by the security guard, the nurse employed in a doctor's office in the centre continued to take up valuable parking space, parking at the curb only twenty feet from the doctor's office instead of in the parking area designated for tenants and personnel. The nurse told the centre manager and the council of merchants that it was none of their concern. The landlord advised the doctor by letter that if further violations occurred, he would have to leave the centre. The doctor spoke to his nurse. She continued parking at the curb.

Outcome: When she had received three more warnings, the centre manager sent a letter to the doctor requesting the premises at the end of the month. The doctor asked if he could stay if he dismissed his nurse, and was told he could. He dismissed her reluctantly. He said she was a good nurse.

Comment: The promotion director said it thereafter became the unwritten rule in that centre that any employee who accumulated three parking warnings (I did not think to ask over what period of time) would be dismissed.

CASE 5: THE CASE OF THE MANIPULATED MERCHANTS

Informant: Shopping-centre manager.

Facts: The manager of a major store in the centre, while not a director on the merchants' council, was active on its advertising committee. He proposed at a weekly meeting that all the merchants hold a five-day flower show on the mall, during which they could put out 'throwaways' advertising the specials they would run during that week only. The other merchants liked the idea and began to develop the plan. Just before the flower show was scheduled to open, a small merchant learned from a newspaper space salesman that the major was coming out with a big newspaper section at the same time as the show. The small merchant went to the centre manager with this information. The centre manager called the large merchant, who confirmed it. On the day that the 'throwaways' on the flower show were put out announcing the individual merchants' specials, the big store announced a major, week-long sales event that overshadowed the merchandising of the rest of the stores in the centre.

Outcome: Everybody talked about the incident, but it was never brought into the open. The consensus was that the big store had conceived the idea of the flower show only to augment its own major promotional effort with advertising moneys and specials put out by the other stores. All the merchants felt bad about it. They recognised there had been no lease violation and that all the stores were free to spend their own money for promotion as they liked, but they objected that the major had not told them about his plans so they could have geared their activity to his. Fortunately, everybody did a good business that week.

While the incident was never brought up in any kind of meeting, the informant said an understanding was arrived at with the merchants that the centre management would assume responsibility in the future for knowing what promotions were being

planned by the major stores, so that this could not happen again. Asked how he could do this, the informant said he and his staff ask the major merchants what they are planning, when they are coming out with circulars, and so forth.

'Do you ask them this in meetings?'

'No, just when we see them. We make it our business to keep informed about their advertising and promotion plans.' The manager is able to do this because he is continually in touch with his tenants. He reported a Rotary Club organised entirely within the centre: fifty-two merchants meet together every week.

'We're very close here. It's just like a little town. Now, at lunch today I talked to seven of my tenants. . . .'

Notes

1. L. Pospisil, 'The Kapauku Papuans and their Law', *Yale University Publications in Anthropology*, vol. 54 (Yale U.P.), p. 274.

2. 'Many ethnographers assume that a given society has a single legal system. They either neglect legal phenomena on the subgroup levels or project these phenomena into the top society level and make them consistent with it. Instead of accepting this smoothed-out picture of a single legal system in a society, the writer suggests recognition of the fact that there are as many such systems as there are functional groups. The legal systems of families, clans and communities, for example, form a hierarchy of what we may call legal levels, according to the inclusiveness of the respective groups.'

QUEUE CULTURE: THE WAITING LINE AS A SOCIAL SYSTEM

Leon Mann

THIS paper deals with the kinds of formal and informal arrangements made in queues to regulate behaviour, to recognise the priority of early-comers, and to inhibit the growth of conflict. The study of queues is an attempt to describe the patterns of behaviour which attend waiting in lines where a great deal of time, discomfort and risk of disappointment must be endured before the desired commodity is obtained. Our focus is on the forms of etiquette developed to regulate life in the queue and to minimise the amount of suffering experienced while waiting. The queue culture provides direction on such matters as place-keeping privileges, sanctions against pushing in and rights of temporary absence from the waiting line. Evidence that the queue constitutes an embryonic social system is based primarily on the study of long, overnight lines for football tickets in Melbourne, Australia.

THE STUDY

Every Saturday afternoon in the month of September, over 100,000 spectators crowd into a stadium in Melbourne, Australia, to watch the 'world series' of Australian rules football.

On 15 August 1967, approximately 10,000 people formed twenty-two queues outside the Melbourne Football Stadium to buy 14,000 sets of tickets for the four games. It was the last opportunity to get tickets because mail application for the bulk of the tickets had been oversubscribed weeks before. A great many of the 10,000 faced disappointment because most queuers usually buy the full allotment of two adult and two children's tickets.

From 6 a.m. until 8 a.m., when the selling windows opened, a team of nine research assistants, male psychology majors from the University of Melbourne, conducted short, standard interviews with 216 people in ten of the twenty-two queues. Each interviewer was randomly assigned a queue. Starting with the first person in line, the procedure was to approach every tenth person. The request was brief and informal: 'I am from the University, and we are doing a study of how people feel about the queues. Would you care to answer a few questions?' Only two refusals were encountered; with the exception of one queue, all interviews were completed by 8 a.m. when the lines began to move. Questions covered attitudes towards the system of queuing, evidence of pushing in and place-keeping, arrangements to make the task of queuing more pleasant as well as estimates of position in line and chances of getting a ticket. Interviewers also made notes on their observations of the physical shape of the queues and on their impressions of the mood and morale of the people in them. Data from the Melbourne stadium are the main source of evidence cited in this paper.

Members of the research team also conducted interviews and made observations in the club queues at suburban Collingwood, Carlton and Richmond, each of which had allocations of 1,000 tickets for club members. Data gathered from the club queues provide additional evidence presented in this paper.

THE QUEUE TRADITION

The system of selling seat bookings for the football finals several weeks before the start of the series was first introduced in 1956. Before 1956, people queued outside the stadium on the day of the game, and 'first in' took the best seats. Over the years, because of the large increase in the number of football followers, the queue system became accepted as the only workable method for selling tickets. Although there were complaints from the public, the great overnight queues became a regular event at the end of a Melbourne winter. And as the queues took on an institutional character, increasing numbers of veterans began to regard them as a kind of cherished tradition or ritual. For example, even during the regular season, although it was possible to get choice seats two hours before the commencement of most Saturday games, long queues formed outside stadiums on Friday, perhaps to train for the big one in August.

The queue of 1965 was perhaps the most remarkable, for in that year 25,000 people waited for 12,500 tickets, some of them for over a week, in mud and drizzling rain. Queuers erected a shanty-town of tents and caravans outside the stadium, and conditions, according to the Melbourne town clerk, rapidly became 'squalid and unhygienic'. In 1966, to prevent a recurrence of the shanty-town, the Melbourne City Council banned tents and camping equipment from the queues and prohibited the lighting of fires. Also, queues or assemblies were not allowed outside the stadium until twenty hours before ticket sales started. The city council regulations made the wait for tickets colder, but much shorter, and accordingly it was decided to retain them the following year.

In the 1967 queues, our interviewers noted that people improvised tents by tying tarpaulins to the side of barricades, and brought stretchers, sleeping bags and supplies of liquor to make themselves comfortable during the wait. Even after a cold night in the open, 26 per cent of the respondents claimed they were happy with the queue system. Only the aged and those who had to go straight to work felt very unhappy about their night out. In 1966, when a sample of 122 queuers were interviewed on a mild afternoon before the ticket windows opened, 47 per cent reported satisfaction with this method of selling tickets.

At Collingwood, the Melbourne City Council regulations did not apply, and accommodations in the first part of the queue resembled a refugee camp. The first three families in line, numbering approximately thirty men, women and children, pitched a bedouin tent on the sidewalk fronting the ticket box, and settled down to a six-day wait around a blazing camp fire. Some enthusiasts moved out of their homes and took up formal residence in the queue. Five days before tickets went on sale, the general secretary of Collingwood, Gordon Carlyon, received a letter addressed to 'Mr Alfred McDougall, c/o Queue outside Collingwood Football Ground, Collingwood, 3066'. The *Melbourne Herald* of 8 August 1967 reported that Mr Carlyon threaded his way through beds and tents on the sidewalk outside the stadium to deliver the letter. Melbournians had not only started to tolerate queues but actually seemed to be enjoying them. One woman outside the Melbourne stadium was heard to remark: 'People are always knocking queues, what I would like to know is what people like myself would do without them' (*Melbourne Age*, 16 August 1967).

It seems that the means behaviour, that is, lining up to get tickets for the event, almost becomes an end in itself, with its own intrinsic rewards and satisfactions. What does queuing mean, and why has it become an important occasion in the lives of these people? The answer lies partly in the publicity and recognition given to the queuers and partly in the challenge and excitement. For several days in August the attention of Melbourne and its mass media is focused on the brave queues outside its stadium. To be able to claim, in football-mad Melbourne, that one has stood through the night and obtained tickets earns the kind of kudos and respect that must have been given to those who fought at Agincourt. And there are other pleasures. Outside the stadium something of a carnival atmosphere prevails. The devotees sing, sip warm drinks, play cards and huddle together around the big charcoal braziers. If he has come as part of a large group, or a cheer squad, the aficionado enjoys a brief taste of communal living and the chance to discuss and debate endlessly the fine points of the game. Above all, football fans regard the great queue as an adventure, an unusual and yet traditional diversion at the end of a Melbourne winter, as the football season approaches its exciting climax.

PROFILE OF THE QUEUER

The typical queuer is male, not yet twenty-five years of age, lives in a working-class suburb, and probably has absented himself from work to wait in line. Together with three friends, he has waited for at least fifteen hours to get tickets to watch his club play in the finals. He cannot explain why he likes football, but he has followed his team faithfully since childhood. He claims he would still be queuing even if his team were not playing, but the scarcity of supporters of non-finals teams in the queue indicates that this is not likely. He has not counted the number of people ahead of him and has no real idea of the number of tickets for sale to the queue. He is fairly confident that he will get tickets, and he does not seem very unhappy about the queue system.

THE PROFESSIONAL QUEUER

When demand exceeds supply, it is inevitable that ticket speculators move into the queue in search of supplies for the flourishing market in hard-to-get tickets.

The Australian football queue contains two kinds of speculators: groups of highly organised people, hired at a fee to wait and buy tickets for large business concerns; and small-time operators, who resell their two tickets to the highest bidder. Often the speculators are university students, whose earnings help to pay tuition fees. Two days before ticket sales opened at the Melbourne stadium, twenty students flew from Tasmania to the mainland to join the queue. The airline company, which hired them to buy the tickets, also provided free return flights, accommodation at a leading hotel and taxis during their stay (*Melbourne Herald*, 12 August 1967). An advertising agency engaged Melbourne University students to stand in line for one dollar an hour each, the tickets to be given away as prizes. Other students, operating as freelance scalpers, asked an outraged public fifty dollars for $5.60 tickets, and had no difficulty getting their price.[1]

It is difficult to estimate the number of speculators in the football queues, as most people would be reluctant to admit to this kind of activity, but it was apparent from the number of advertisements for tickets in the 'Wanted to Sell' columns of the Melbourne newspapers that a large proportion of queuers turned professional in the week following the ticket sales.

In the Melbourne stadium lines, very few people actually counted their position, perhaps because they believed there was no point in it since there was no accurate information available about the number of tickets available to each line. In club queues, however, a different set of conditions obtained. There was usually a single mammoth queue (at Richmond it included over 3000 people and ringed the perimeter of the stadium). More important, the number of tickets on sale was well publicised, and therefore it was possible to make a fairly accurate estimate of the chances of getting tickets, if the person had accurate information about his position in line. Accordingly, the estimates of position in the club queues were somewhat more accurate than at the stadium because people either had taken the trouble to count the number ahead or had consulted with a queue 'counter'. Queue counters are boys who count the queue at regular intervals, if it is long and winding. Queue counters, like ticket speculators, 'invent' businesses to go along with what began simply as a necessary social act. At Carlton a group of boys went backwards and forwards during the night, counting the queue, and at Richmond the counters turned professional and, for a fee (ten cents), gave each customer up-to-date information on the number of people ahead and behind, as well as topical news and gossip.

THE PRE-QUEUE

A queue is a line of persons waiting in turn to be served, according to order of arrival. But the act of queuing involves more than the acquisition of a right to prior service because of early arrival. To validate this priority, the person must also spend time in the queue, not only to show latecomers that he occupies a given position, but also to demonstrate that his right to priority is confirmed by an unquestionable willingness to undergo further suffering to get the commodity.

If all that is required to reserve a place in a queue is the act of registering order of arrival, everyone would make an effort to be present at the time of queue formation. This would either lead to uncontrolled competition and hostility at the time of registration,

or more probably the formation of pre-queues to establish recognition of the right to priority in the official queue.

The pre-queue is an unofficial line which forms spontaneously before the official, recognised queue is allowed to form. The Melbourne football queues were not allowed to form until 3 p.m. of the afternoon before the sale of tickets. To enforce this regulation, police erected a perimeter of barricades around the wall of the stadium. Nevertheless, hundreds of people gathered in the park hours before queuing was officially allowed to start and, without police direction or intervention, spontaneously formed lines outside the barricades. At 3 p.m., when the barricades were removed, they folded their chairs, and keeping the lines intact, filed in perfect order to the ticket windows to commence the official seventeen-hour wait. The formation of a pre-queue, in this instance, almost certainly functioned to prevent an explosive situation which could have occurred had people failed to sort themselves into some kind of recognised order before the official line started. The lack of competition for positions among the early-comers can be explained in terms of the reward-cost structure in the first part of the line. There is little to be gained from being first, rather than twentieth or fiftieth (all are virtually guaranteed a ticket), but there is much to be lost if aggressive competition leads to physical damage and general disorder.

SERVING TIME IN QUEUES

There is a curious dilemma in the overnight queue. If there is a unanimous willingness to respect the order of arrival, it is pointless to require everyone to spend an uncomfortable night in the open. But if large numbers absent themselves, those remaining to protect the queue from outsiders will feel that their greater inputs of time and suffering now outweigh the merits of early arrival and entitle them to priority of service. Also, they will feel no responsibility for minding the places of people who, by their absence, are in no position to offer reciprocal place-minding services. In recognition of the conflicting considerations of unnecessary suffering caused by continuous occupancy, and the necessity to validate one's position by spending some time in residence, various arrangements are made which function to lessen the ordeal while protecting the rights of early-comers. Usually the arrangements represent a compromise which allows the queuer to take brief leaves of absence while retaining undisputed rights of re-entry.

In Australian football lines, 'time-out' is accomplished by two informal arrangements. Early-comers, who usually come in groups of four or five, often organise a 'shift system', in which members spend one hour on with four hours off. One person can hold up to four places until the relief reports back to take over as group custodian. In our survey, an average of 39 per cent of respondents in the first 100 of every queue reported that they had organised a shift system; in the latter part of the queue only 24 per cent reported participation in a shift system. Sometimes the system involves a large group of people who share not only place-keeping duties but also facilities for eating, sleeping and entertainment. The *Melbourne Herald* of 15 August 1967 described a seventeen-year-old girl, one of twelve people who took turns to leave the queue to eat and sleep in one of the few trailers found outside the Melbourne stadium. The same newspaper carried a story of a young scalper who combined business with pleasure: 'I was one of a group of 20 students who stood together all night in the queue outside the ground. We were well organised. A couple of us kept our positions, and the others went out on the town' (*Melbourne Herald*, 22 August 1967).

It is rare for queuers at the head of the line to come alone; 94 per cent of the respondents questioned in the first 100 of every line reported that they had come with others (see Table 26.1). How-

TABLE 26.1

PATTERNS OF QUEUE-RELATED BEHAVIOUR AS A FUNCTION OF POSITION IN LINE: MELBOURNE FOOTBALL QUEUES, 1967

Actual position in line	*Member of an organised shift system*		*Came with others*		*Observed hostile response to push-in attempt*		*Consider place-keeping permissible*	
10–100 (N=95)	39%	Z=1·91, P<0·05	94%	Z=2·81, $<P$0·01	46%	N.S.	25%	N.S.
110–200 (N=82)	23%	N.S.	78%	Z=1·41, NN.S.	33%	N.S.	35%	N.S.
210–330 (N=39)	25%		64%		27%	C	24%	

Note: Number of respondents varies slightly for each question.

ever, a large minority towards the end of each line came alone; while their need for time-outs was less pressing, they also made arrangements to cover brief absences if necessary. It is an accepted practice to 'stake a claim' in a queue by leaving some item of personal property. One can keep a place in a line with a labelled box, folding chair, haversack or sleeping bag for quite long periods. The object stands for the person and his place, symbolism reminiscent of burial customs of the ancient Egyptians.[2] During the early hours of waiting, when many people were enjoying a carefree game of football in the surrounding park, the queues often consisted of one part people to two parts inanimate objects. The norm in leaving position markers is that one must not be absent for periods longer than two to three hours. In the Collingwood queue of 1966, irate latecomers, who noticed that many people in the middle of the queue had not made an appearance for most of the day, spontaneously seized their boxes and burnt them. The latecomers were protesting the violation of the principle of serving time to earn occupancy of a position. In the ensuing mêlée, scores of people made significant advances in their positions. Because arrangements for absence from the football queue are of necessity extremely informal, inefficiency and abuse often occur. To ensure protection of their valued positions, some do not trust the shift or market systems but prefer to keep a constant vigil which lasts the entire life of the queue.

First come, first served, the fundamental concept of queueing, is a basic principle of the behaviour referred to as distributive justice.[3] There is a direct correspondence between inputs (time spent waiting) and outcomes (preferential service). Generally, if a person is willing to invest large amounts of time and suffering in an activity, people who believe there should be an appropriate fit between effort and reward will respect his right to priority. We have seen, however, that the principle of distributive justice is elaborated to encompass the need for leaves of absence in marathon queues. In recognition of the fact that continuous residence in the line imposes great hardship, members come to an agreement on the minimum inputs of time necessary to validate occupancy of a position. It is reasonable to claim that rules regulating time spent in and out of the line are the essential core of the queue culture.[4]

QUEUE JUMPING

Place-keeping and pushing in violate the principle of first come, first served. When people at the end of a queue feel certain that the violation does not jeopardise their own chances of obtaining the commodity, there is likely to be some irritation but no attempt to eject the offender. Stronger measures are likely, however, if people at the tail end believe that the lengthening of the line worsens their prospects of receiving service.

Since there is a great deal at stake, football queuers are especially annoyed by any attempt to jump the queue, and they adopt a variety of physical and social techniques to keep people in line. At certain times in the life of the queue when police supervision was minimal, queuers had to devise their own constraints. The most extreme constraint was physical force. During the early hours of 15 August, five men were taken to hospital after four separate brawls broke out in the ticket lines (*Melbourne Herald*, 15 August 1967). The strategic placements of barriers acts as a constraint against would-be infiltrators. It was observed that people in the middle of the queue worked together to erect barricades from material left in the park. Keeping close inter-person distance also serves to maintain the 'territory' in the face of would-be intruders. At times of maximum danger, and in the hour before the ticket windows opened, there was a visible bunching together, or shrinkage, in the physical length of the queue, literally a closing of the ranks. The exercise of effective social constraints depends on the capacity for cohesive action on the part of the queuers. At the stadium, whenever outsiders approached the head of the queue, they were intimidated by vociferous catcalls and jeering. Ordinarily, this mode of protecting the queue was successful during daylight, the pressure of concerted disapproval inhibiting all but the boldest. During the hours of darkness, social pressure proved less effective; the knowledge that one cannot be seen easily undermines social pressure and shaming as a technique.

Despite these constraints, many latecomers attempted to push in, and it was apparent that some succeeded. Letters to the newspapers by disappointed queuers testified to the activity of queue jumpers. One man who missed out in the Carlton queue claimed that he had been dislodged from 185th to 375th place in the span of two hours. When asked, 'Has anyone tried to push in?' respond-

ents in every part of each queue reported that they had witnessed attempts to jump the queue, but only in a minority of cases had the intruder been ejected. According to the reports of our respondents, the act of intrusion was usually met with passivity rather than a physically hostile response, especially towards the end of the line, where people came along and were not organised for dealing with intruders (see Table 26.1). Yet, when asked what they would do if someone tried to crash the queue immediately in front of them, respondents were almost unanimous in claiming that they would resort to physical force.

According to our respondents' reports, pushing in occurred most often near the tail of the queue. This seems puzzling at first, for, if someone is going to risk pushing in, it seems sensible to try at the front, where there is a greater certainty of getting tickets. However, we must bear in mind the more effective policing at the front, as well as the decreased risk towards the rear of the queue, where, in absolute numbers, fewer people are put out by the violation and hence there is less likelihood of concerted action. In brief, opposition to a queue jumper decreases as a function of the number of people whose chances of getting tickets are affected by the intrusion; at the end of the line there are very few such people. Ironically, however, it is these people, regardless of where pushing in occurs, who stand to lose most by the infraction because their chances of getting tickets are put in even greater jeopardy.

Why does the queue fail to act in unison to dismiss the queue jumper? To some extent, the varying interests of people in different parts of the queue provide an answer. People at the front of the queue do not care particularly about pushing in which occurs behind them because they do not suffer from the intrusion. Of course, if queue jumping becomes widespread, the early-comers show concern because their positions may be threatened by late-comers who realise that the entire line is vulnerable. But usually they have nothing to gain and much to lose from becoming involved in policing the queue. It is surprising, however, that people after the point of intrusion do not act together to expel the violator, since they all suffer equally by the loss of a place. It seems that responsibility for evicting a trespasser falls squarely on the shoulders of the person who is the immediate victim of the violation, that is, the person directly behind the violator. Those further back may jeer and catcall, but the immediate victim is expected to take the initiative in ejecting the queue jumper. The reasoning seems to be that the victim, either through his passive looks or careless surveillance of his territory, must have given some encouragement to the queue jumper, so he is now obliged to handle the situation without causing unpleasantness for other people.

The reluctance of queuers to exert physical action against queue jumpers may also be related to the nature of informal versus formal organisation of the queue. In any informal queue, there are many signs of organisational control, role prerogatives and orderly behaviour, which are almost exactly the same as those in well-organised queues, where there is real policing and monitoring of the line. Therefore, people will assume that the informal queue will function in much the same way as the organised queue. When acts of pushing in and disorder occur, members of the queue realise they were mistaken and jeer spontaneously, exert informal pressure and make threats to preserve their positions. If verbal constraints fail, physical violence emerges as a last resort. At this point there is a reluctance to pursue the matter further because more may be lost from physical action than from a small loss in position. The person who jumps the queue could be desperate, and the immediate victim anticipates the possibility that a struggle could cause injury and damage. If the police action is unsuccessful, the person is made to look foolish in the eyes of the onlookers. It is also possible that, if the struggle sets off a widespread mêlée, he stands to lose more than face and position. Therefore, if verbal censure fails, members of the queue fall back on a conspiracy of silence to ignore minor violations. Resorting to physical violence seems to represent a kind of public acknowledgement that the queue is no longer organised and under control. Once this happens, a grave danger exists that people in less favourable positions, as well as outsiders, will take advantage of what is then recognised to be a helpless, unorganised queue. To prevent this, occasional minor infractions, if they are not met successfully by verbal threats and jeering, are seldom handled by physical threats and violence. The use of physical methods, especially if they prove unsuccessful, is a signal to others that the queue organisation is about to disintegrate completely, and this may actually serve to encourage an epidemic of queue jumping.

One reason for the prevalence of pushing in, and the failure to

exert effective action against it, is the confusion which exists between illegal acts of entry and the somewhat more acceptable act of place-keeping. Because place-keeping occurs fairly frequently, it is not always clear whether an individual who moves boldly into a line is attempting to crash the queue or is merely joining his group. Therefore many are reluctant to challenge the entry of outsiders during the early hours of the queue. Although the custom of place-keeping is a cause of friction, only informal rules have been formulated to regulate its practice. Of the respondents, 29 per cent believed that it is permissible to keep a place for someone, and that people behind would not care (see Table 26.1). However, only a handful of queuers admitted to actually keeping a place for someone. People do not admit freely to place-keeping because the newcomer usually makes his appearance only in the last hour before tickets go on sale, and people already in line are likely to be very resentful.

THE QUEUE AS A SOCIAL SYSTEM

The queue, although made up of numerous groups of strangers gathered together temporarily, emerges as an embryonic social system with a set of norms for controlling conflict.

Parsons[5] maintains that social systems develop spontaneously whenever two or more people come into some stabilised, patterned mode of interaction. He lists three properties of any social system: (1) two or more actors occupying differentiated statuses or positions and performing differentiated roles; (2) some organised pattern governing the relationships of the members, describing their rights and obligations with respect to one another; and (3) some set of common norms and values, together with various types of shared cultural objects and symbols.

The long, overnight queue has all three characteristics of a social system. While the queue may not directly allocate different statuses or roles to its members, the members themselves assume different roles. In and around every queue there is a host of people: professional and hired speculators, queue counters, custodians, vigilantes, police and officials, performing a variety of queue-related tasks.

The other two properties of a social system – an organised pattern of relationships and a set of common norms – are readily identifiable in the queue. Order of arrival governs the relationship among members, while the shift system and the practice of time-outs controls the network of rights and obligations. Moreover, there are shared norms about the desirability of distributive justice, as reflected in the set of rules regulating place-keeping and pushing in.

Interactive systems, such as the queue, develop within the matrix of a long-established socio-cultural system which defines roles, normative standards and goals. When a large number of people gather together and priority of service has value, a line is formed. All members bring to the new queue a host of ideas about the roles they should play, and develop firm notions about the way in which deviant behaviour should be punished. Roles in the queue are drawn from and are moulded by the institutional system of the larger society. The precise form of social organisation, the sharing and division of labour inherent in the shift system, the preferred modes of policing the queue, the development of businesses and ticket speculation, the notion that one must earn one's place in line by spending time in it, even the very reason for queueing itself, reflect the character of the surrounding society.

The culture of the queue also draws upon and incorporates elements in the broader culture. The importance of time as a value in Western society is reflected in the emphasis placed on serving time, and restrictions on time-outs. The way in which people orient themselves towards a scarce commodity, their preference for co-operation, the entrepreneurial zeal they display in scalping tickets and charging fees for counting the queue, is a function of broad culture patterns, as well as the way society has taught them to behave.

The queue, moreover, is subject to sanctioning pressure from outside officials and onlookers, who try to bring it into conformity with societal expectations. Ultimately, of course, each queue has to work out its own final set of mutual adjustments in which socially prescribed rules about queueing are modified and embellished in various ways. A prime example of these adjustments is to be found in the various interpretations of the rules governing leaves of absence from the line.

While the queue system is embedded in a larger social matrix, it is also composed of many subsystems – groups, cliques and co-acting

individuals – whose physical presence reinforces the very idea or concept of a line.

According to Parsons and Smelser,[6] there are four functional problems, or imperatives, faced by every social system: goal attainment, adaptation, integration and latency. The queue, even though it is a relatively minor, short-lived social system, must confront these four problems.

The problem of goal attainment is to keep the system moving steadily towards the collective goal of its members, in this case, the purchase of tickets in an orderly manner with a minimum of unpleasantness. Before the system can move towards this goal, however, a host of instrumental and technical problems must be solved. Many of the problems are external to the queue, in the sense that they are not under the control of its members; for example, the seller (not the customer) must decide when and where people should be allowed to queue, how many lines should be formed, how many tickets will be made available, what limitation will be placed on the number of tickets sold to each customer, and so on. But the question of how to begin the queue, especially if people have gathered before the official starting time, is a problem left to the members themselves. As we have seen, the football fans solved this difficulty by forming a pre-queue, which became the officially recognised queue when the barriers to the ticket boxes were lowered.

Adaptation is the problem of bringing facilities and resources to the system which enable it to come to terms with the environment. One aspect of adaptation is the active manipulation of the environment. Thus queuers formed their line along concrete paths, constructed barriers out of material found in the park, erected shelters by tying tarpaulins to the barriers, built fires and even brought in trailers, to make their temporary living quarters as comfortable as possible.

Manipulation of the system itself to blend with the environment is another aspect of adaptation. For the most part, the queue, rather than a single file of people, consisted of numerous knots of people, two and three abreast, who sat side by side to facilitate efficient communication and social interaction.

The integrative problem, perhaps the most distinctive in any social system, is concerned with the maintenance of appropriate emotional and social ties among members of the system. In order to achieve its goals, the system must establish and maintain a high degree of solidarity and cohesion. In the queue, cohesion is achieved by establishing informal rules which are kept sufficiently general to allow individual members to adjust to the normative pattern. Those who stay out too long, and therefore are unable to make the line viable, are sanctioned, or lose their place. In a sense this represents a form of turn taking; if the queue structure is to be preserved, only some members can be permitted to take leaves of absence at any one time.

The group or clique, by means of the shift system, regulates turn taking for its members, and this ensures continuity of the line. The group, since it is the carrier of queue culture, brings a high level of solidarity to the line.

At the head of the line, the group takes on the characteristics of a community. Large family groups share eating, rest and recreational facilities, and time spent together serves to strengthen the feelings of community. It is likely that the major factor underlying the effective policing of the head and middle regions of the line is the presence of large, co-ordinated groups. The breakdown of defence against intruders in the end part of every line can be attributed primarily to the fragmented, isolated nature of the membership.

But even in parts of the line where organised groups are less prominent, individuals trade on a mutual trust which allows them to ask one another to 'mind my place' and feel confident that they will be vouched for when they return from a brief leave of absence.

The latency function is reflected in two related but different problems – pattern maintenance and tension management.

Pattern maintenance is the problem faced by an individual in reconciling the conflicting norms and demands imposed by his participation in the queue. Many members experience role conflicts arising out of their obligations to the queue and to their family or work roles. As we have seen, some queuers solve the problem by moving their entire family into the line. Others are faced with a different kind of dilemma: whether or not to keep a place for friends. The member who fails to commit himself to the queue norms is subject to considerable social pressure. If rules governing leaves of absence are not observed, the member is likely to find himself no longer part of the queue.

Tension management is the related problem of maintaining a

level of commitment sufficient to perform the required role. To cope with tension and fatigue, members introduce a variety of entertainments in and around the line, such as pick-up games, story and joke telling and beer parties. Time-out from the queue is, however, the major mode of tension management.

The queue system is mostly concerned with the problem of pattern maintenance and tension management because these are the most significant from the viewpoint of continuous participation and control. Of course, at critical times in the life of the queue, the other three functional problems require attention. Indeed, all four must be solved if the system is to continue in the state of equilibrium necessary for control and order.

CULTURAL VALUES AND QUEUEING

How do the rules of etiquette develop in queues, and how do people come to recognise and respect them? To some extent, the answer lies in attitudes towards queues in general, which reflect broad cultural differences in sensitivity to queue norms, as well as the individuals' history of experience in waiting lines. Behaviour in queues is a function of many variables: kind and length of queue; the importance of inputs and the value of the commodity; cultural and subcultural differences in respect for time, order and the rights of others; and individual differences in such personality characteristics as aggressiveness and assertiveness.

The anthropologist Hall has suggested that a cultural value of egalitarianism is responsible for the manner in which queues and queueing are treated with deference in Western society. In *The Silent Language*, Hall writes:

> As a general rule, whenever services are involved we feel that people should queue up in order of arrival. This reflects the basic equalitarianism of our culture. In cultures where a class system or its remnants exist, such ordinality may not exist. That is, where society assigns rank for certain purposes, or whenever ranking is involved, the handling of space will reflect this. To us it is regarded as a democratic virtue for people to be served without reference to the rank they hold in their occupational group. The rich and poor alike are accorded equal opportunity to buy and be waited upon in the order of arrival. In a line at the theatre Mrs Gottrocks is no better than anyone else.[7]

While there is merit in the thesis that a strain of equality in a culture is related to the orderliness of public behaviour, it would seem that egalitarianism alone does not explain respect for priority in queues. Consider the English, a people famous for their strictly democratic queueing behaviour. It is safe to say that the English have a more rigid social structure and are more class-conscious than Americans, and yet their public behaviour is probably more orderly. The relationship between cultural equality and public orderliness is attenuated in the area of queueing because waiting in line is not a habit of all social classes in Western society. It is reasonable to suppose that if Mrs Gottrocks joined a theatre or a football line in the United States, Australia or England, she would not be treated differently than anyone else, but it would be a rare event for someone of Mrs Gottrocks's status to use a line. Ordinarily, in both class-conscious and relatively class-free societies, the privileged classes circumvent the line altogether and get their tickets through agents or other contacts. Our point, then, is that queueing is confined largely to the less-privileged groups in society. It might be more accurate to speak of a subcultural value of equalitarianism in public behaviour.

In his recent book, *The Hidden Dimension*, Hall[8] continues his examination of the basis of public behaviour and makes the assertion that respect for queues can also be attributed to a cultural value of orderliness. Presumably, culture patterns in different societies emerge on a continuum of order–disorder, and, like the value of equality, a strain of orderliness runs through life in a society.

One difficulty with the orderliness hypothesis is that the principle of queueing is not always maintained in situations where there are no officials or police present to scrutinise and control the outbursts of unruly behaviour. Only if the principle is supported religiously under these conditions, can one speak of a culturally ingrained disposition towards orderliness.

In brief, to seek an explanation for organised co-operative behaviour in queues, we cannot appeal only to cultural values of egalitarianism and orderliness. Answers must also be sought with reference to the queue as a social system in which these values

often receive only token or symbolic acknowledgement. The main concern of the queue membership is maintenance of a state of control and order while the system moves towards its goal. Indeed, values of orderliness, egalitarianism, and distributive justice are invoked to help maintain that order, but they are primarily used and modified to make life in the queue tolerable, rather than obeyed in a rigid and mechanical manner.

CONCLUSION

This paper described how patterned regularities in behaviour and attitudes emerge to regulate life in an overnight football queue. Although arrangements made to control behaviour in the queue are informal, they are clearly identifiable, and it is appropriate to regard them as constituting a kind of culture. The queue, which possesses the characteristics of a social system, attempts to solve the set of functional problems confronted by every social system.

Our major findings were: (*a*) the growth of a queue tradition in which large numbers of people return annually to share the experience of waiting for tickets overnight in the open; (*b*) an increasing professionalisation of the queue, marked by an influx of speculators and middlemen who profit by the increased demand for tickets; (*c*) the formation of unofficial pre-queues to recognise the priority of people who arrived before the start of the official queue; (*d*) elaboration of the principle of first come, first served to control the amount of time spent in and out of the queue ('shift' and 'marker' systems, which control 'time-outs', were developed to regulate leaves of absence from the line); (*e*) social constraints, and less often physical constraints, used to control queue jumping and to govern the practice of place-keeping.

It is appropriate to conclude that queueing behaviour, a neglected area of social research, could be a rich source of ideas for students of crowd behaviour, judgemental processes, cross-national differences, and the influence of cultural values on public behaviour.

Notes

1. Speculation in the physical position itself is not found in Australian queues, as it is in waiting lines for Broadway hit shows. At smash hits it is not unusual for people to make a business of getting in line early in order to sell their advanced positions to late-comers for a large fee (see *Life*, 24 September 1956).
2. Markers such as notebooks, coats, newspapers and umbrellas are often used to defend a 'reserved' space in public places, such as a crowded cafeteria or study hall. R. Sommer, *Personal Space* (New York, 1969).
3. G. C. Homans, *Social Behavior: Its Elementary Forms* (New York, 1961).
4. Queue systems with inbuilt guarantees of distributive justice are to be found in both the United States and U.S.S.R. At the weekly line for tickets to the Metropolitan Opera in New York, an unofficial 'keeper of the list' registers applicants in order of arrival, assigns numbers, and checks names when the queuers appear for roll call every three hours (see *New Yorker*, 14 January 1967). In Moscow, when scarce goods go on sale, a series of queue custodians take turns 'standing guard' and list the names of interested customers as they arrive throughout the night. I. R. Levine, *Main Street U.S.S.R.* (New York, 1959).
5. T. Parsons, *The Social System* (New York, 1951).
6. Idem and N. J. Smelser, *Economy and Society* (New York, 1957).
7. E. T. Hall, *The Silent Language* (New York, 1959) p. 157.
8. Idem, *The Hidden Dimension* (New York, 1964).

The student of government and politics is interested in structures and institutions in societies with and without the state, but he is also concerned with political attitudes. Attitudes towards authority and obedience, the state and rulers are all part of a political culture that has a direct bearing on how people behave as voters, legislators, civil servants or the ruled.

Political culture, dealing as it does with men's ideas, beliefs and values about politics and government, covers a very wide field, and this is well brought out in three extracts which follow. R. H. Tawney uses a broad historical canvas in his classical study of puritanism, 'Religion and the Rise of Capitalism'. In this book, Tawney sets out to examine the part which the great revolution in religious ideas played in shaping economic and political attitudes and behaviour in European society during the sixteenth, seventeenth and eighteenth centuries. A narrower canvas is used in the piece by M. Weiner, 'Struggle Against Power: Notes on Indian Political Behaviour'. From Tawney's study we have jumped in terms of geography and culture over a vast distance. But the same basic point stands out – that men's attitudes and beliefs play an important role in shaping their political behaviour.

The final article of this group carries this point further by examining political culture in a modern technologically advanced society. Lewis Lipsitz's 'Work Life and Political Attitudes: A Study of Manual Workers' examines one specific aspect of our industrial society – attitudes of manual labourers in factory conditions.

To a consideration of the structure and institutions of government in society, therefore, there must be added an account of the political culture if an understanding is to be gained of government in society. At the same time, if each feature is not to be viewed simply as something distinct, there needs to be a conception of the whole of which the separate features are a part. The idea of a political system as expounded by Easton is one general theory that has been developed to encompass the different distinct aspects. He places institutions like legislatures, political parties or interest groups and other political activities and attitudes in their context with a political system, and thus allows realistic comparisons to be made between the political systems in different societies. In recent years a number of attempts have been made to develop general theory in order to integrate particular knowledge: Easton's is only one such theory, but it is a particularly influential theory and one that has been considerably developed.

AN APPROACH TO THE ANALYSIS OF POLITICAL SYSTEMS

David Easton

I. SOME ATTRIBUTES OF POLITICAL SYSTEMS

In an earlier work I have argued for the need to develop general, empirically oriented theory as the most economical way in the long run to understand political life. Here I propose to indicate a point of view that, at the least, might serve as a springboard for discussion of alternative approaches and, at most, as a small step in the direction of a general political theory. I wish to stress that what I have to say is a mere orientation to the problem of theory; outside of economics and perhaps psychology, it would be presumptuous to call very much in social science 'theory', in the strict sense of the term.

Furthermore, I shall offer only a Gestalt of my point of view, so that it will be possible to evaluate, in the light of the whole, those parts that I do stress. In doing this, I know I run the definite risk that the meaning and implications of this point of view may be only superficially communicated; but it is a risk I shall have to undertake since I do not know how to avoid it sensibly.

The study of politics is concerned with understanding how authoritative decisions are made and executed for a society. We can try to understand political life by viewing each of its aspects piecemeal. We can examine the operation of such institutions as political parties, interest groups, government and voting; we can study the nature and consequences of such political practices as manipulation, propaganda and violence; we can seek to reveal the structure within which these practices occur. By combining the results we can obtain a rough picture of what happens in any self-contained political unit.

In combining these results, however, there is already implicit the notion that each part of the larger political canvas does not stand alone but is related to each other part; or, to put it positively, that the operation of no one part can be fully understood without reference to the way in which the whole itself operates. I have suggested in my book, *The Political System*,[1] that it is valuable to adopt this implicit assumption as an articulate premise for research and

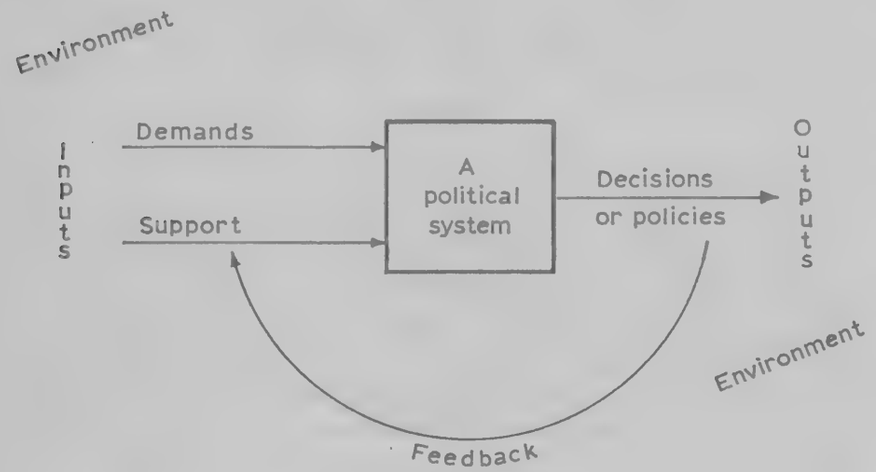


Fig. 27.1

to view political life as a system of interrelated activities. These activities derive their relatedness or systemic ties from the fact that they all more or less influence the way in which authoritative decisions are formulated and executed for a society.

Once we begin to speak of political life as a system of activity, certain consequences follow for the way in which we can undertake to analyse the working of a system. The very idea of a system suggests that we can separate political life from the rest of social activity, at least for analytical purposes, and examine it as though for the moment it were a self-contained entity surrounded by, but clearly distinguishable from, the environment or setting in which it operates. In much the same way, astronomers consider the solar system a complex of events isolated for certain purposes from the rest of the universe.

Furthermore, if we hold the system of political actions as a unit before our mind's eye, as it were, we can see that what keeps the system going are inputs of various kinds. These inputs are converted by the processes of the system into outputs and these, in turn, have consequences both for the system and for the environment in which the system exists. The formula here is very simple but, as I hope to show, also very illuminating: inputs – political system or processes – outputs. These relationships are shown diagrammatically in Fig. 27.1. This diagram represents a very primitive 'model' – to dignify it with a fashionable name – for approaching the study of political life.

Political systems have certain properties because they are systems. To present an overall view of the whole approach, let me identify the major attributes, say a little about each, and then treat one of these properties at somewhat greater length, even though still inadequately.

1. *Properties of identification.* To distinguish a political system from other social systems, we must be able to identify it by describing its fundamental units and establishing the boundaries that demarcate it from units outside the system.

(*a*) *Units of a political system.* The units are the elements of which we say a system is composed. In the case of a political system, they are political actions. Normally it is useful to look at these as they structure themselves in political roles and political groups.

(*b*) *Boundaries.* Some of the most significant questions with regard to the operation of political systems can be answered only if we bear in mind the obvious fact that a system does not exist in a vacuum. It is always immersed in a specific setting or environment. The way in which a system works will be in part a function of its response to the total social, biological and physical environment.

The special problem with which we are confronted is how to distinguish systematically between a political system and its setting. Does it even make sense to say that a political system has a boundary dividing it from its setting? If so, how are we to identify the line of demarcation?

Without pausing to argue the matter, I would suggest that it is useful to conceive of a political system as having a boundary in the same sense as a physical system. The boundary of a political system is defined by all those actions more or less directly related

to the making of binding decisions for a society; every social action that does not partake of this characteristic will be excluded from the system and thereby will automatically be viewed as an external variable in the environment.

2. *Inputs and outputs.* Presumably, if we select political systems for special study, we do so because we believe that they have characteristically important consequences for society, namely, authoritative decisions. These consequences I shall call the outputs. If we judged that political systems did not have important outputs for society, we would probably not be interested in them.

Unless a system is approaching a state of entropy – and we can assume that this is not true of most political systems – it must have continuing inputs to keep it going. Without inputs the system can do no work; without outputs we cannot identify the work done by the system. The specific research tasks in this connection would be to identify the inputs and the forces that shape and change them, to trace the process through which they are transformed into outputs, to describe the general conditions under which such processes can be maintained, and to establish the relationship between outputs and succeeding inputs of the system.

From this point of view, much light can be shed on the working of a political system if we take into account the fact that much of what happens within a system has its birth in the efforts of the members of the system to cope with the changing environment. We can appreciate this point if we consider a familiar biological system such as the human organism. It is subject to constant stress from its surroundings to which it must adapt in one way or another if it is not to be completely destroyed. In part, of course, the way in which the body works represents responses to needs that are generated by the very organisation of its anatomy and functions; but in large part, in order to understand both the structure and the working of the body, we must also be very sensitive to the inputs from the environment.

In the same way, the behaviour of every political system is to some degree imposed upon it by the kind of system it is, that is, by its own structure and internal needs. But its behaviour also reflects the strains occasioned by the specific setting within which the system operates. It may be argued that most of the significant changes within a political system have their origin in shifts among the external variables. Since I shall be devoting the bulk of this article to examining some of the problems related to the exchange between political systems and their environments, I shall move on to a rapid description of other properties of political systems.

3. *Differentiation within a system.* As we shall see in a moment, from the environment come both energy to activate a system and information with regard to which the system uses this energy. In this way a system is able to do work. It has some sort of output that is different from the input that enters from the environment. We can take it as a useful hypothesis that if a political system is to perform some work for anything but a limited interval of time, a minimal amount of differentiation in its structure must occur. In fact, empirically it is impossible to find a significant political system in which the same units all perform the same activities at the same time. The members of a system engage in at least some minimal division of labour that provides a structure within which action takes place.

4. *Integration of a system.* This fact of differentiation opens up a major area of inquiry with regard to political systems. Structural differentiation sets in motion forces that are potentially disintegrative in their results for the system. If two or more units are performing different kinds of activity at the same time, how are these activities to be brought into the minimal degree of articulation necessary if the members of the system are not to end up in utter disorganisation with regard to the production of the outputs of interest to us? We can hypothesise that if a structured system is to maintain itself, it must provide mechanisms whereby its members are integrated or induced to co-operate in some minimal degree so that they can make authoritative decisions.

II. INPUTS: DEMANDS

Now that I have mentioned some major attributes of political systems that I suggest require special attention if we are to develop a generalised approach, I want to consider in greater detail the way in which an examination of inputs and outputs will shed some light on the working of these systems.

Among inputs of a political system there are two basic kinds: demands and support. These inputs give a political system its dynamic character. They furnish it both with the raw material or

information that the system is called upon to process and with the energy to keep it going.

The reason why a political system emerges in a society at all – that is, why men engage in political activity – is that demands are being made by persons or groups in the society that cannot all be fully satisfied. In all societies one fact dominates political life: scarcity prevails with regard to most of the valued things. Some of the claims for these relatively scarce things never find their way into the political system but are satisfied through the private negotiations of or settlements by the persons involved. Demands for prestige may find satisfaction through the status relations of society; claims for wealth are met in part through the economic system; aspirations for power find expression in educational, fraternal, labour and similar private organisations. Only where wants require some special organised effort on the part of society to settle them authoritatively may we say that they have become inputs of the political system.

Systematic research would require us to address ourselves to several key questions with regard to these demands.

1. How do demands arise and assume their particular character in a society? In answer to this question, we can point out that demands have their birth in two sectors of experience: either in the environment of a system or within the system itself. We shall call these the external and internal demands, respectively.

Let us look at the external demands first. I find it useful to see the environment not as an undifferentiated mass of events but rather as systems clearly distinguishable from one another and from the political system. In the environment we have such systems as the ecology, economy, culture, personality, social structure and demography. Each of these constitutes a major set of variables in the setting that helps to shape the kind of demands entering a political system. For purposes of illustrating what I mean, I shall say a few words about culture.

The members of every society act within the framework of an on-going culture that shapes their general goals, specific objectives and the procedures that the members feel ought to be used. Every culture derives part of its unique quality from the fact that it emphasises one or more special aspects of behaviour and this strategic emphasis serves to differentiate it from other cultures with respect to the demands that it generates. As far as the mass of the people is concerned, some cultures, such as our own, are weighted heavily on the side of economic wants, success, privacy, leisure activity and rational efficiency. Others, such as that of the Fox Indians, strive towards the maintenance of harmony, even if in the process the goals of efficiency and rationality may be sacrificed. Still others, such as the Kachins of highland Burma, stress the pursuit of power and prestige. The culture embodies the standards of value in a society and thereby marks out areas of potential conflict, if the valued things are in short supply relative to demand. The typical demands that will find their way into the political process will concern the matters in conflict that are labelled important by the culture. For this reason we cannot hope to understand the nature of the demands presenting themselves for political settlement unless we are ready to explore systematically and intensively their connection with the culture. And what I have said about culture applies, with suitable modifications, to other parts of the setting of a political system.

But not all demands originate or have their major locus in the environment. Important types stem from situations occurring within a political system itself. Typically, in every on-going system, demands may emerge for alterations in the political relationships of the members themselves, as the result of dissatisfaction stemming from these relationships. For example, in a political system based upon representation, in which equal representation is an important political norm, demands may arise for equalising representation between urban and rural voting districts. Similarly, demands for changes in the process of recruitment of formal political leaders, for modifications of the way in which constitutions are amended, and the like may all be internally inspired demands.

I find it useful and necessary to distinguish these from external demands because they are, strictly speaking, not inputs of the system but something that we can call 'withinputs', if we can tolerate a cumbersome neologism, and because their consequences for the character of a political system are more direct than in the case of external demands. Furthermore, if we were not aware of this difference in classes of demands, we might search in vain for an explanation of the emergence of a given set of internal demands if we turned only to the environment.

2. How are demands transformed into issues? What determines

whether a demand becomes a matter for serious political discussion or remains something to be resolved privately among the members of society? The occurrence of a demand, whether internal or external, does not thereby automatically convert it into a political *issue*. Many demands die at birth or linger on with the support of an insignificant fraction of the society and are never raised to the level of possible political decision. Others become issues, an issue being a demand that the members of a political system are prepared to deal with as a significant item for discussion through the recognised channels in the system.

The distinction between demands and issues raises a number of questions about which we need data if we are to understand the processes through which claims typically become transformed into issues. For example, we would need to know something about the relationship between a demand and the location of its initiators or supporters in the power structures of the society, the importance of secrecy as compared with publicity in presenting demands, the matter of timing of demands, the possession of political skills or know-how, access to channels of communication, the attitudes and states of mind of possible publics, and the images held by the initiators of demands with regard to the way in which things get done in the particular political system. Answers to matters such as these would possibly yield a conversion index reflecting the probability of a set of demands being converted into live political issues.

If we assume that political science is primarily concerned with the way in which authoritative decisions are made for a society, demands require special attention as a major type of input of political systems. I have suggested that demands influence the behaviour of a system in a number of ways. They constitute a significant part of the material upon which the system operates. They are also one of the sources of change in political systems, since as the environment fluctuates it generates new types of demand-inputs for the system. Accordingly, without this attention to the origin and determinants of demands we would be at a loss to be able to treat rigorously not only the operation of a system at a moment of time but also its change over a specified interval. Both the statics and historical dynamics of a political system depend upon a detailed understanding of demands, particularly of the impact of the setting on them.

III. INPUTS: SUPPORT

Inputs of demands alone are not enough to keep a political system operating. They are only the raw material out of which finished products called decisions are manufactured. Energy in the form of actions or orientations promoting and resisting a political system, the demands arising in it, and the decisions issuing from it must also be put into the system to keep it running. This input I shall call support.[2] Without support, demands could not be satisfied or conflicts in goals composed. If demands are to be acted upon, the members of a system undertaking to pilot the demands through to their transformation into binding decisions and those who seek to influence the relevant processes in any way must be able to count on support from others in the system. Just how much support, from how many and which members of a political system, are separate and important questions that I shall touch on shortly.

What do we mean by support? We can say that A supports B either when A acts on behalf of or when he orients himself favourably towards B's goals, interests and actions. Supportive behaviour may thus be of two kinds. It may consist of actions promoting the goals, interests and actions of another person. We may vote for a political candidate, or defend a decision by the highest court of the land. In these cases, support manifests itself through overt action.

On the other hand, supportive behaviour may involve not external observable acts, but those internal forms of behaviour we call orientations or states of mind. As I use the phrase, a supportive state of mind is a deep-seated set of attitudes or predispositions, or a readiness to act on behalf of some other person. It exists when we say that a man is loyal to his party, attached to democracy, or infused with patriotism. What such phrases as these have in common is the fact that they refer to a state of feelings on the part of a person. No overt action is involved at this level of description, although the implication is that the individual will pursue a course of action consistent with his attitudes. Where the anticipated action does not flow from our perception of the state of mind, we assume that we have not penetrated deeply enough into the true feelings of the person but have merely skimmed off his surface attitudes.

Supportive states of mind are vital inputs for the operation and maintenance of a political system. For example, it is often said that the struggle in the international sphere concerns mastery over men's minds. To a certain extent this is true. If the members of a political system are deeply attached to a system or its ideals, the likelihood of their participating in either domestic or foreign politics in such a way as to undermine the system is reduced by a large factor. Presumably, even in the face of considerable provocation, ingrained supportive feelings of loyalty may be expected to prevail.

We shall need to identify the typical mechanisms through which supportive attitudes are inculcated and continuously reinforced within a political system. But our prior task is to specify and examine the political objects in relation to which support is extended.

1. *The domain of support*

Support is fed into the political system in relation to three objects: the community, the regime and the government. There must be convergence of attitude and opinion as well as some willingness to act with regard to each of these objects. Let us examine each in turn.

(*a*) *The political community.* No political system can continue to operate unless its members are willing to support the existence of a group that seeks to settle differences or promote decisions through peaceful action in common. The point is so obvious – being dealt with usually under the heading of the growth of national unity – that it may well be overlooked; and yet it is a premise upon which the continuation of any political system depends. To refer to this phenomenon we can speak of the political community. At this level of support we are not concerned with whether a government exists or whether there is loyalty to a constitutional order. For the moment we only ask whether the members of the group that we are examining are sufficiently oriented towards each other to want to contribute their collective energies towards pacific settlement of their varying demands.

The American Civil War is a concrete illustration of the cessation of input of support for the political community. The war itself was definitive evidence that the members of the American political system could no longer contribute to the existence of a state of affairs in which peaceful solution of conflicting demands was the rule. Matters had come to the point where it was no longer a question of whether the South would support one or another alternative government, or whether it could envision its demands being satisfied through the normal constitutional procedures. The issue turned on whether there was sufficient mutual identification among the members of the system for them to be able to work together as a political community. Thus in any political system, to the extent that there is an in-group or we-group feeling and to the extent that the members of the system identify one another as part of this unit and exclude others according to some commonly accepted criteria, such as territoriality, kinship or citizenship, we shall say that they are putting in support for the political community.

(*b*) *The regime.* Support for a second major part of a political system helps to supply the energy to keep the system running. This aspect of the system I shall call the regime. It consists of all those arrangements that regulate the way in which the demands put into the system are settled and the way in which decisions are put into effect. They are the so-called rules of the game, in the light of which actions by members of the system are legitimated and accepted by the bulk of the members as authoritative. Unless there is a minimum convergence of attitudes in support of these fundamental rules – the constitutional principles, as we call them in Western society – there would be insufficient harmony in the actions of the members of a system to meet the problems generated by their support of a political community. The fact of trying to settle demands in common means that there must be known principles governing the way in which resolutions of differences of claims are to take place.

(*c*) *The government.* If a political system is going to be able to handle the conflicting demands put into it, not only must the members of the system be prepared to support the settlement of these conflicts in common and possess some consensus with regard to the rules governing the mode of settlement; they must also be ready to support a government as it undertakes the concrete tasks involved in negotiating such settlements. When we come to the outputs of a system, we shall see the rewards that are available to a government for mobilising support. At this point, I just wish to draw attention to this need on the part of a government for

support if it is going to be able to make decisions with regard to demands. Of course, a government may elicit support in many ways: through persuasion, consent or manipulation. It may also impose unsupported settlements of demands through threats of force. But it is a familiar axiom of political science that a government based upon force alone is not long for this world; it must buttress its position by inducing a favourable state of mind in its subjects through fair or foul means.

The fact that support directed to a political system can be broken down conceptually into three elements – support for the community, regime and government – does not mean, of course, that in the concrete case support for each of these three objects is independent. In fact we might and normally do find all three kinds of support very closely intertwined, so that the presence of one is a function of the presence of one or both of the other types.

For example, withdrawal of support from the government of Louis XVI in effect also meant that members of the French monarchical system were challenging at least the regime; as it turned out in the ensuing revolution and civil war, there was even doubt whether the members of the system would continue to support a unified political community. In this case, what was initially opposition to the ruling sovereign – that is, to the government – quickly turned out to signify a lack of sufficient support for the regime and ultimately, to some extent, for the political community. But this is not always so and fortunately, from the point of view of social order, it is not typically the case. We are accustomed to calling for a change of government without thereby suggesting dissatisfaction with the regime or community. And at times, although this is less frequently true, the community shows sufficient intention to continue as a co-operating group to be able to accept a challenge to the regime. From 1832 to the 1880s England underwent a serious modification in its regime, introducing the basic elements of a system of popular democracy, without serious diminution of input of support at the community level. It is always a matter for empirical inquiry to discover the degree to which support at any one level is dependent upon support at the others.

This very brief discussion of support points up one major fact. If a system is to absorb a variety of demands and negotiate some sort of settlement among them, it is not enough for the members of the system to support only their own demands and the particular government that will undertake to promote these demands. For the demands to be processed into outputs it is equally essential that the members of the system stand ready to support the existence of a political community and some stable rules of common action that we call the regime.

2. *Quantity and scope of support*

How much support needs to be put into a system and how many of its members need to contribute such support if the system is to be able to do the job of converting demands to decisions? No ready answer can be offered. The actual situation in each case would determine the amount and scope required. We can, however, visualise a number of situations that will be helpful in directing our attention to possible generalisations.

Under certain circumstances very few members need to support a system at any level. The members might be dull and apathetic, indifferent to the general operations of the system, its progress or decisions. In a loosely connected system such as India has had, this might well be the state of mind of by far the largest segment of the membership. Either in fact they have not been affected by national decisions or they have not perceived that they were so affected. They may have little sense of identification with the present regime and government and yet, with regard to the input of demands, the system may be able to act on the basis of the support offered by the known 3 per cent of the Western-oriented politicians and intellectuals who are politically active. In other words, we can have a small minority putting in quantitatively sufficient supportive energy to keep the system going. However, we can venture the hypothesis that where members of a system are putting in numerous demands, there is a strong probability that they will actively offer support or hostility at one of the three levels of the system, depending upon the degree to which these demands are being met through appropriate decisions.

Alternatively, we may find that all the members of a system are putting in support, but the amount may be so low as to place one or all aspects of the system in jeopardy. Modern France is perhaps a classic illustration.[3] The input of support at the level of the political community is probably adequate for the maintenance of France as a national political unit. But for a variety of historical

and contemporary reasons, there is considerable doubt as to whether the members of the French political system are putting in anything but a low order of support to the regime or any particular government. This low amount of support, even though spread over a relatively large segment of the population, leaves the French political system on somewhat less secure foundations than is the case with India. There support is less widespread but more active – that is, quantitatively greater – on the part of a minority. As this illustration indicates, the amount of support is not necessarily proportional to its scope.

It may seem from the above discussion as though the members of a political system either put in support or withhold it – that is demonstrate hostility or apathy. In fact, members may and normally do simultaneously engage in supportive and hostile behaviour. What we must be interested in is the net balance of support.

IV. MECHANISMS OF SUPPORT

To this point I have suggested that no political system can yield the important outputs we call authoritative decisions unless, in addition to demands, support finds its way into the system. I have discussed the possible object to which support may be directed, and some problems with regard to the domain, quantity and scope of support. We are now ready to turn to the main question raised by our attention to support as a crucial input: how do systems typically manage to maintain a steady flow of support? Without it a system will not absorb sufficient energy from its members to be able to convert demands to decisions.

In theory, there might be an infinite variety of means through which members could be induced to support a system; in practice, certain well-established classes of mechanisms are used. Research in this area needs to be directed to exploring the precise way in which a particular system utilises these mechanisms and to refining our understanding of the way in which they contribute to the making of authoritative policy.

A society generates support for a political system in two ways: through outputs that meet the demands of the members of society; and through the processes of politicisation. Let us look at outputs first.

1. *Outputs as a mechanism of support*

An output of a political system, it will be recalled, is a political decision or policy. One of the major ways of strengthening the ties of the members to their system is through providing decisions that tend to satisfy the day-to-day demands of these members. Fundamentally this is the truth that lies in the aphorism that one can fool some of the people some of the time but not all of them all of the time. Without some minimal satisfaction of demands, the ardour of all but the most fanatical patriot is sure to cool. The outputs, consisting of political decisions, constitute a body of specific inducements for the members of a system to support that system.

Inducements of this kind may be positive or negative. Where negative, they threaten the members of the system with various kinds of sanctions ranging from a small monetary fine to physical detention, ostracism or loss of life, as in our own system with regard to the case of legally defined treason. In every system support stems in part from fear of sanctions or compulsion; in autocratic systems the proportion of coerced support is at a maximum. For want of space I shall confine myself to those cases where positive incentives loom largest.

Since the specific outputs of a system are policy decisions, it is upon the government that the final responsibility falls for matching or balancing outputs of decisions against input of demand. But it is clear that to obtain the support of the members of a system through positive incentives, a government need not meet all the demands of even its most influential and ardent supporters. Most governments, or groups such as political parties that seek to control governments, succeed in building up a reserve of support. This reserve will carry the government along even though it offends its followers, so long as over the extended short run these followers perceive the particular government as one that is in general favourable to their interests. One form that this reserve support takes in Western society is that of party loyalty, since the party is the typical instrument in a mass industrialised society for mobilising and maintaining support for a government. However, continuous lack of specific rewards through policy decisions ultimately leads to the danger that even the deepest party loyalty may be shaken.

For example, labour has continued to support the Democratic

Party even though much of the legislation promoted by members of that party has not served to meet labour's demands. In some measure, large sections of labour may continue to vote and campaign vigorously on behalf of the Democratic Party because they have no realistic alternative other than to support this party; but in addition the Democrats have built up in recent years, especially during the Roosevelt era, a considerable body of goodwill. It would take repeated neglect of labour's demands on the part of the Democratic Party to undermine the strong urban working-class support directed towards it and the government that the party dominates from time to time.

Thus a system need not meet *all the demands* of its members so long as it has stored up a reserve of support over the years. Nor need it satisfy even *some of the demands* of all its members. Just whose demands a system must seek to meet, how much of their demands, at what time, and under what conditions are questions for special research. We can say in advance that at least the demands of the most influential members require satisfaction. But this tells us little unless we know how to discover the influentials in a political system and how new sets of members rise to positions of influence.[4]

The critical significance of the decisions of governments for the support of the other two aspects of a system – namely, the political community and the regime – is clear from what I have said above. Not all withdrawal of support from a government has consequences for the success or failure of a regime or community. But persistent inability of a government to produce satisfactory outputs for the members of a system may well lead to demands for changing of the regime or for dissolution of the political community. It is for this reason that the input–output balance is a vital mechanism in the life of a political system.

2. *Politicisation as a mechanism of support*

It would be wrong to consider that the level of support available to a system is a function exclusively of the outputs in the form of either sanctions or rewards. If we did so conclude, we could scarcely account for the maintenance of numerous political systems in which satisfaction of demands has been manifestly low, in which public coercion is limited, and yet which have endured for epochs. Alternately, it might be difficult to explain how political systems could endure and yet manage to flout or thwart urgent demands, failing thereby to render sufficient *quid pro quo* for the input of support. The fact is that whatever reserve of support has been accumulated through past decisions is increased and reinforced by a complicated method for steadily manufacturing support through what I shall call the process of politicisation. It is an awkward term, but nevertheless an appropriately descriptive one.

As each person grows up in a society, through a network of rewards and punishments the other members of society communicate to and instil in him the various institutionalised goals and norms of that society. This is well known in social research as the process of socialisation. Through its operation a person learns to play his various social roles. Part of these goals and norms relate to what the society considers desirable in political life. The ways in which these political patterns are learned by the members of society constitute what I call the process of politicisation. Through it a person learns to play his political roles, which include the absorption of the proper political attitudes.

Let us examine a little more closely something of what happens during the process of politicisation. As members of a society mature, they must absorb the various orientations towards political matters that one is expected to have in that society. If the expectations of the members of society with regard to the way each should behave in specific political situations diverged beyond a certain range, it would be impossible to get common action with regard to the making of binding decisions. It is essential for the viability of an orderly political system that the members of the system have some common basic expectations with regard to the standards that are to be used in making political evaluations, to the way people will feel about various political matters, and to the way members of the system will perceive and interpret political phenomena.

The mechanism through which this learning takes place is of considerable significance in understanding how a political system generates and accumulates a strong reserve of support. Although we cannot pursue the details, we can mention a few of the relevant dimensions. In the first place, of course, the learning or politicisation process does not stop at any particular period for the individual; it starts with the child and, in the light of our knowledge of learning, may have its deepest impact through the teen age.

The study of the political experiences of and the influences operating on the child and the adolescent emerges as an important and neglected area of research.

In the second place, the actual process of politicisation at its most general level brings into operation a complex network of rewards and punishments. For adopting the correct political attitudes and performing the right political acts, for conforming to the generally accepted interpretations of political goals, and for undertaking the institutionalised obligations of a member of the given system, we are variously rewarded or punished. For conforming we are made to feel worthy, wanted and respected and often obtain material advantages such as wealth, influence, improved opportunities. For deviating beyond the permissible range, we are made to feel unworthy, rejected, dishonoured and often suffer material losses.

This does not mean that the pattern of rewards and punishments is by any means always effective; if it were, we would never have changed from the Stone Age. A measure of non-conformity may at certain stages in the life history of a political system itself become a respected norm. Even where this is not the case, the most seductive rewards and the severest punishments will never succeed in preventing some of the members of a system from pursuing what they consider to be their inextinguishable interests and from seeking, with varying degrees of success, to change the goals and norms of the system. This is one of the important sources of political change closely associated with changes in the inputs of demands that are due to a changing environment. But we cannot pursue this crucial matter of the nature of political change, as it would lead us off in a new direction.

In the third place, the means used for communicating the goals and norms to others tend to be repetitive in all societies. The various political myths, doctrines and philosophies transmit to each generation a particular interpretation of the goals and norms. The decisive links in this chain of transmission are parents, siblings, peers, teachers, organisations and social leaders, as well as physical symbols such as flags or totems, ceremonies and rituals freighted with political meaning.

These processes through which attachments to a political system become built into the maturing member of a society I have lumped together under the rubric of politicisation. They illustrate the way in which members of a system learn what is expected of them in political life and how they ought to do what is expected of them. In this way they acquire knowledge about their political roles and a desire to perform them. In stable systems the support that accrues through these means adds to the reservoir of support being accumulated on a day-to-day basis through the outputs of decisions.[5] The support obtained through politicisation tends to be relatively – although, as we have seen, not wholly – independent of the vagaries of day-to-day outputs.

When the basic political attachments become deeply rooted or institutionalised, we say that the system has become accepted as legitimate. Politicisation therefore effectively sums up the way in which legitimacy is created and transmitted in a political system. And it is an empirical observation that in those instances where political systems have survived the longest, support has been nourished by an ingrained belief in the legitimacy of the relevant governments and regimes.

What I am suggesting here is that support resting on a sense of the legitimacy of a government and regime provides a necessary reserve if the system is to weather those frequent storms when the more obvious outputs of the system seem to impose greater hardships than rewards. Answers to questions concerning the formation, maintenance, transmission and change of standards of legitimacy will contribute generously to an understanding of the way in which support is sufficiently institutionalised so that a system may regularly and without excessive expenditure of effort transform inputs of demand into outputs of decisions.

That there is a need for general theory in the study of political life is apparent. The only question is how best to proceed. There is no one royal road that can be said to be either the correct one or the best. It is only a matter of what appears at the given level of available knowledge to be the most useful. At this stage it appears that system theory, with its sensitivity to the input–output exchange between a system and its setting, offers a fruitful approach. It is an economical way of organising presently disconnected political data and promises interesting dividends.

Notes

1. (New York, 1953).

2. The concept support has been used by Talcott Parsons in an unpublished paper entitled 'Reflections on the Two-Party System'. I

am pleased to note that in this article Professor Parsons also seems to be moving in the direction of input–output analysis of political problems, although the extent to which he uses other aspects of system theory is not clear to me.

3. [This article first published in 1957 (Ed.).]

4. See C. W. Mills, *The Power Elite* (New York, 1956).

5. In primitive systems, politicisation, not outputs of decisions, is normally the chief mechanism.

RELIGION AND THE RISE OF CAPITALISM

R. H. Tawney

(trying to see life through the eyes of a Puritan)

His life is that of a soldier in hostile territory. He suffers in spirit the perils which the first settlers in America endured in body, the sea behind, the untamed desert in front, a cloud of inhuman enemies on either hand. Where Catholic and Anglican had caught a glimpse of the invisible, hovering like a consecration over the gross world of sense, and touching its muddy vesture with the unearthly gleam of a divine, yet familiar, beauty, the Puritan mourned for a lost Paradise and a creation sunk in sin. Where they had seen society as a mystical body, compact of members varying in order and degree, but dignified by participation in the common life of Christendom, he saw a bleak antithesis between the spirit which quickeneth and an alien, indifferent or hostile world. Where they had reverenced the decent order whereby past was knit to present, and man to man, and man to God, through fellowship in works of charity, in festival and fast, in the prayers and ceremonies of the Church, he turned with horror from the filthy rags of human righteousness. Where they, in short, had found comfort in a sacrament, he started back from a snare set to entrap his soul.

> We receive but what we give,
> And in our life alone does Nature live.

Too often, contemning the external order as unspiritual, he made it, and ultimately himself, less spiritual by reason of his contempt.

Those who seek God in isolation from their fellow-men, unless trebly armed for the perils of the quest, are apt to find, not God, but a devil, whose countenance bears an embarrassing resemblance to their own. The moral self-sufficiency of the Puritan nerved his will, but it corroded his sense of social solidarity. For, if each individual's destiny hangs on a private transaction between himself and his Maker, what room is left for human intervention? A servant of Jehovah more than of Christ, he revered God as a Judge rather than loved him as a Father, and was moved less by compassion for his erring brethren, than by impatient indignation at the blindness of vessels of wrath who 'sinned their mercies'. A spiritual aristocrat, who sacrificed fraternity to liberty, he drew from his idealisation of personal responsibility a theory of individual rights, which, secularised and generalised, was to be among the most potent explosives that the world has known. He drew from it also a scale of ethical values, in which the traditional scheme of Christian virtues was almost exactly reversed, and which, since he was above all things practical, he carried as a dynamic into the routine of business and political life.

For, since conduct and action, though availing nothing to attain the free gift of salvation, are a proof that the gift has been accorded, what is rejected as a means is resumed as a consequence, and the Puritan flings himself into practical activities with the demonic energy of one who, all doubts allayed, is conscious that he is a sealed and chosen vessel. Once engaged in affairs, he brings to them both the qualities and limitations of his creed, in all their remorseless logic. Called by God to labour in his vineyard, he has within himself a principle at once of energy and of order, which makes himself irresistible both in war and in the struggles of commerce. Convinced that character is all and circumstances nothing, he sees in the poverty of those who fall by the way, not a misfortune to be pitied and relieved, but a moral failing to be condemned, and in riches, not an object of suspicion – though like other gifts they may be abused – but the blessing which rewards the triumph of energy and will. Tempered by self-examination, self-discipline, self-control, he is the practical ascetic, whose victories are won not in the cloister, but on the battlefield, in the counting-house, and in the market.

This temper, of course with infinite varieties of quality and emphasis, found its social organ in those middle and commercial classes who were the citadel of the Puritan spirit, and whom, 'ennobled by their own industry and virtue', Milton described as the

standard-bearers of progress and enlightenment. We are so accustomed to think of England as *par excellence* the pioneer of economic progress, that we are apt to forget how recently that role has been assumed. In the Middle Ages it belonged to the Italians, in the sixteenth century to the Netherland dominions of the Spanish Empire, in the seventeenth to the United Provinces and, above all, to the Dutch.

The England of Shakespeare and Bacon was still largely medieval in its economic organisation and social outlook, more interested in maintaining customary standards of consumption than in accumulating capital for future production, with an aristocracy contemptuous of the economic virtues, a peasantry farming for subsistence amid the organised confusion of the open-field village, and a small, if growing, body of jealously conservative craftsmen. In such a society Puritanism worked like the yeast which sets the whole mass fermenting. It went through its slack and loosely knot texture like a troop of Cromwell's Ironsides through the disorderly cavalry of Rupert. Where, as in Ireland, the elements were so alien that assimilation was out of the question, the result was a wound that festered for three centuries. In England the effect was that at once of an irritant and of a tonic. Puritanism had its own standards of social conduct, derived partly from the obvious interests of the commercial classes, partly from its conception of the nature of God and the destiny of man. These standards were in sharp antithesis, both to the considerable surviving elements of feudalism in English society, and to the policy of the authoritarian State, with its ideal of an ordered and graded society, whose different members were to be maintained in their traditional status by the pressure and protection of a paternal monarchy. Sapping the former by its influence, and overthrowing the latter by direct attack, Puritanism became a potent force in preparing the way for the commercial civilisation which finally triumphed at the Revolution. . . .

* * *

For it was not merely as the exponent of certain tenets as to theology and church government, but as the champion of interests and opinions embracing every side of the life of society, that the Puritan movement came into collision with the Crown. In reality, as is the case with most heroic ideologies, the social and religious aspects of Puritanism were not disentangled; they presented themselves, both to supporters and opponents, as different facets of a single scheme. 'All that crossed the views of the needy courtiers, the proud encroaching priests, the thievish projectors, the lewd nobility and gentry . . . whoever could endure a sermon, modest habit or conversation, or anything good – all these were Puritans.' The clash was not one of theories – a systematic and theoretical individualism did not develop till after the Restoration – but of contradictory economic interests and incompatible conceptions of social expediency.

The economic policy haltingly pursued by the Government of Charles I bore some resemblance to the system of which a more uncompromising version was developed between 1661 and 1685 by Colbert in France. It was one which favoured an artificial and State-promoted capitalism – a capitalism resting on the grant of privileges and concessions to company promoters who would pay for them, and accompanied by an elaborate system of State control, which again, if partly inspired by a genuine solicitude for the public interest, was too often smeared with an odious trail of finance. It found its characteristic expression in the grant of patents, in the revival of the royal monopoly of exchange business, against which the City had fought under Elizabeth, in attempts to enforce by administrative action compliance with the elaborate and impracticable code controlling the textile trades and to put down speculation in foodstuffs, and in raids on enclosing landlords, on employers who paid in truck or evaded the rates fixed by assessment, and on justices who were negligent in the administration of the Poor Laws. Such measures were combined with occasional plunges into even more grandiose schemes for the establishment of county granaries, for taking certain industries into the hands of the Crown, and even for the virtual nationalisation of the cloth manufacture.

'The very genius of that nation of people,' wrote Strafford to Laud of the Puritans, 'leads them always to oppose, as well civilly as ecclesiastically, all that ever authority ordains for them.' Against this whole attempt to convert economic activity into an instrument of profit for the Government and its hangers-on – against, no less, the spasmodic attempts of the State to protect peasants against landlords, craftsmen against merchants and consumers against middlemen – the interests which it thwarted and curbed

revolted with increasing pertinacity. Questions of taxation, on which attention has usually been concentrated, were in reality merely one element in a quarrel, which had its deeper cause in the collision of incompatible social philosophies. The Puritan tradesman had seen his business ruined by a monopoly granted to a needy courtier, and cursed Laud and his Popish soap. The Puritan goldsmith or financier had found his trade as a bullion-broker hampered by the re-establishment of the ancient office of Royal Exchanger, and secured a resolution from the House of Commons, declaring that the patent vesting it in Lord Holland, and the proclamation forbidding the exchanging of gold and silver by unauthorised persons, were a grievance. The Puritan moneylender had been punished by the Court of High Commission, and railed at the interference of bishops in temporal affairs. The Puritan clothier, who had suffered many things at the hands of interfering busybodies, dispatched from Whitehall to teach him his business, averted discreet eyes when the Wiltshire workmen threw a more than usually obnoxious Royal Commissioner into the Avon, and, when the Civil War came, rallied to the Parliament. The Puritan country gentleman had been harried by Depopulation Commissions, and took his revenge with the meeting of the Long Parliament. The Puritan merchant had seen the Crown both squeeze money out of his company and threaten its monopoly by encouraging courtly interlopers to infringe its charter. The Puritan Member of Parliament had invested in colonial enterprises, and had ideas as to commercial policy which were not those of the Government. Confident in their own energy and acumen, proud of their success, and regarding with profound distrust the interference both of Church and of State with matters of business and property rights, the commercial classes, in spite of their attachment to a militant mercantilism in matters of trade, were, even before the Civil War, more than half converted to the administrative nihilism which was to be the rule of social policy in the century following it. Their demand was the one which is usual in such circumstances. It was that business affairs should be left to be settled by businessmen, unhampered by the intrusions of an antiquated morality or by misconceived arguments of public policy.

The separation of economic from ethical interests, which was the note of all this movement, was in sharp opposition to religious tradition, and it did not establish itself without a struggle. Even in the very capital of European commerce and finance, an embittered controversy was occasioned by the refusal to admit usurers to communion or to confer degrees upon them; it was only after a storm of pamphleteering, in which the theological faculty of the University of Utrecht performed prodigies of zeal and ingenuity, that the States of Holland and West Friesland closed the agitation by declaring that the Church had no concern with questions of banking. In the French Calvinist Churches the decline of discipline had caused lamentations a generation earlier. In America, the theocracy of Massachusetts, merciless alike to religious liberty and to economic licence, was about to be undermined by the rise of new States like Rhode Island and Pennsylvania, whose tolerant, individualist and utilitarian temper was destined to find its greatest representative in the golden common sense of Benjamin Franklin. 'The sin of our too great fondness for trade, to the neglecting of our more valuable interests,' wrote a Scottish divine in 1709, when Glasgow was on the eve of a triumphant outburst of commercial enterprise, 'I humbly think will be written upon our judgement. . . . I am sure the Lord is remarkably frowning upon our trade . . . since it was put in the room of religion.'

In England, the growing disposition to apply exclusively economic standards to social relations evoked from Puritan writers and divines vigorous protests against usurious interest, extortionate prices and the oppression of tenants by landlords. The faithful, it was urged, had interpreted only too literally the doctrine that the sinner was saved, not by works, but by faith. Usury, 'in time of Popery an odious thing', had become a scandal. Professors, by their covetousness, caused the enemies of the reformed religion to blaspheme. The exactions of the forestaller and regrater were never so monstrous or so immune from interference. The hearts of the rich were never so hard, nor the necessities of the poor so neglected. 'The poor able to work are suffered to beg; the impotent, aged and sick are not sufficiently provided for, but almost starved with the allowance of 3*d* and 4*d* a piece a week. . . . These are the last times indeed. Men generally are all for themselves. And some would set up such, having a form of religion, without the power of it.'

These utterances came, however, from that part of the Puritan mind which looked backward. That which looked forward found

in the rapidly growing spirit of economic enterprise something not uncongenial to its own temper, and went out to welcome it as an ally. What in Calvin had been a qualified concession to practical exigencies, appeared in some of his later followers as a frank idealisation of the life of the trader, as the service of God and the training-ground of the soul. Discarding the suspicion of economic motives, which had been as characteristic of the reformers as of medieval theologians, Puritanism in its later phases added a halo of ethical sanctification to the appeal of economic expediency, and offered a moral creed, in which the duties of religion and the calls of business ended their long estrangement in an unanticipated reconciliation. Its spokesmen pointed out, it is true, the peril to the soul involved in a single-minded concentration on economic interests. The enemy, however, was not riches, but the bad habits sometimes associated with them, and its warnings against an excessive preoccupation with the pursuit of gain wore more and more the air of after-thoughts, appended to teaching the main tendency and emphasis of which were little affected by these incidental qualifications. It insisted, in short, that money-making, if not free from spiritual dangers, was not a danger and nothing else, but that it could be, and ought to be, carried on for the greater glory of God.

The conception to which it appealed to bridge the gulf sprang from the very heart of Puritan theology. It was that expressed in the characteristic and oft-used phrase, 'a Calling'. The rational order of the universe is the work of God, and its plan requires that the individual should labour for God's glory. There is a spiritual calling and a temporal calling. It is the first duty of the Christian to know and believe in God; it is by faith that he will be saved. But faith is not a mere profession, such as that of Talkative of Prating Row, whose 'religion is to make a noise'. The only genuine faith is the faith which produces works. 'At the day of Doom men shall be judged according to their fruits. It will not be said then, Did you believe? but, Were you doers, or talkers only?' The second duty of the Christian is to labour in the affairs of practical life, and this second duty is subordinate only to the first. 'God,' wrote a Puritan divine, 'doth call every man and woman . . . to serve him in some peculiar employment in this world, both for their own and the common good. . . . The Great Governor of the world hath appointed to every man his proper post and province, and let him be never so active out of his sphere, he will be at a great loss, if he do not keep his own vineyard and mind his own business.' . . .

* * *

It would be misleading to dwell on the limitations of Puritan ethics without emphasising the enormous contribution of Puritanism to political freedom and social progress. The foundation of democracy is the sense of spiritual independence, which nerves the individual to stand alone against the powers of this world, and in England, where squire and parson, lifting arrogant eyebrows at the insolence of the lower orders, combined to crush popular agitation, as a menace at once to society and to the Church, it is probable that democracy owes more to Nonconformity than to any other single movement. The virtues of enterprise, diligence and thrift are the indispensable foundation of any complex and vigorous civilisation. It was Puritanism which, by investing them with a supernatural sanction, turned them from an unsocial eccentricity into a habit and a religion. Nor would it be difficult to find notable representatives of the Puritan spirit, in whom the personal austerity, which was the noblest aspect of the new ideal, was combined with a profound consciousness of social solidarity which was the noblest aspect of that which it displaced. Firmin the philanthropist, and Bellers the Quaker, whom Owen more than a century later hailed as the father of his doctrines, were pioneers of Poor Law reform. The Society of Friends, in an age when the divorce between religion and social ethics was almost complete, met the prevalent doctrine that it was permissible to take such gain as the market offered, by insisting on the obligation of good conscience and forbearance in economic transactions, and on the duty to make the honourable maintenance of the brother in distress a common charge.

The general climate and character of a country are not altered, however, by the fact that here and there it has peaks which rise into an ampler air. The distinctive note of Puritan teaching was different. It was individual responsibility, not social obligation. Training its pupils to the mastery of others through the mastery of self, it prized as a crown of glory the qualities which arm the spiritual athlete for his solitary contest with a hostile world, and dismissed concern with the social order as the prop of weaklings

and the Capua of the soul. Both the excellences and the defects of that attitude were momentous for the future. It is sometimes suggested that the astonishing outburst of industrial activity, which took place after 1760, created a new type of economic character, as well as a new system of economic organisation. In reality, the ideal which was later to carry all before it, in the person of the inventor and engineer and captain of industry, was well established among Englishmen before the end of the seventeenth century. Among the numerous forces which had gone to form it, some not inconsiderable part may reasonably be ascribed to the emphasis of the life of business enterprise as the appropriate field for Christian endeavour, and on the qualities needed for success in it, which was characteristic of Puritanism. These qualities, and the admiration of them, remained, when the religious reference, and the restraints which it imposed, had weakened or disappeared.

CONCLUSION

> 'There is a certaine man that shortly after my fyrst sermon, beynge asked if he had bene at the sermon that day, answered, yea. I praye you, said he, how lyked you hym? Mary, sayed he, even as I lyked hym alwayes – a sedicious fellow.'
>
> Latimer, *Seven Sermons before King Edward VI*

Societies, like individuals, have their moral crises and their spiritual revolutions. The student can observe the results which these cataclysms produce, but he can hardly without presumption attempt to appraise them, for it is at the fire which they kindled that his own small taper has been lit. The rise of a naturalistic science of society, with all its magnificent promise of fruitful action and of intellectual light; the abdication of the Christian Churches from departments of economic conduct and social theory long claimed as their province; the general acceptance by thinkers of a scale of ethical values, which turned the desire for pecuniary gain from a perilous, if natural, frailty into the idol of philosophers and the mainspring of society – such movements are written large over the history of the tempestuous age which lies between the Reformation and the full light of the eighteenth century. Their consequences have been worked into the very tissue of modern civilisation. Posterity still stands too near their source to discern the ocean into which these streams will flow.

In an historical age the relativity of political doctrines is the tritest of commonplaces. But social psychology continues too often to be discussed in serene indifference to the categories of time and place, and economic interests are still popularly treated as though they formed a kingdom over which the *Zeitgeist* bears no sway. In reality, though inherited dispositions may be constant from generation to generation, the system of valuations, preferences and ideals – the social environment within which individual character functions – is in process of continuous change, and it is in the conception of the place to be assigned to economic interests in the life of society that change has in recent centuries been most comprehensive in its scope, and most sensational in its consequences. The isolation of economic aims as a specialised object of concentrated and systematic effort, the erection of economic criteria into an independent and authoritative standard of social expediency, are phenomena which, though familiar enough in classical antiquity, appear, at least on a grand scale, only at a comparatively recent date in the history of later civilisations. The conflict between the economic outlook of East and West, which impresses the traveller today, finds a parallel in the contrast between medieval and modern economic ideas, which strikes the historian.

The elements which combined to produce that revolution are too numerous to be summarised in any neat formula. But, side by side with the expansion of trade and the rise of new classes to political power, there was a further cause, which, if not the most conspicuous, was not the least fundamental. It was the contraction of the territory within which the writ of religion was conceived to run. The criticism which dismisses the concern of Churches with economic relations and social organisation as a modern innovation finds little support in past history. What requires explanation is not the view that these matters are part of the province of religion, but the view that they are not. When the age of the Reformation begins, economics is still a branch of ethics, and ethics of theology; all human activities are treated as falling within a single scheme, whose character is determined by the spiritual destiny of mankind; the appeal of theorists is to natural law, not to utility; the legitimacy of economic transactions is tried by reference, less to the movements of the market, than to moral standards derived from the traditional teaching of the

Christian Church; the Church itself is regarded as a society wielding theoretical, and sometimes practical, authority in social affairs.

The secularisation of political thought, which was to be the work of the next two centuries, had profound reactions on social speculation, and by the Restoration the whole perspective, at least in England, has been revolutionised. Religion has been converted from the keystone which holds together the social edifice into one department within it, and the idea of a rule of right is replaced by economic expediency as the arbiter of policy and the criterion of conduct. From a spiritual being, who, in order to survive, must devote a reasonable attention to economic interests, man seems sometimes to have become an economic animal, who will be prudent, nevertheless, if he takes due precautions to assure his spiritual well-being.

The result is an attitude which forms so fundamental a part of modern political thought, that both its precarious philosophical basis and the contrast which it offers with the conceptions of earlier generations are commonly forgotten. Its essence is a dualism which regards the secular and the religious aspects of life, not as successive stages within a larger unity, but as parallel and independent provinces, governed by different laws, judged by different standards, and amenable to different authorities. To the most representative minds of the Reformation as of the Middle Ages, a philosophy which treated the transactions of commerce and the institutions of society as indifferent to religion would have appeared, not merely morally reprehensible, but intellectually absurd. Holding as their first assumption that the ultimate social authority is the will of God, and that temporal interests are a transitory episode in the life of spirits which are eternal, they state the rules to which the social conduct of the Christian must conform, and, when circumstances allow, organise the discipline by which those rules may be enforced. By their successors in the eighteenth century the philosophy of Indifferentism, though rarely formulated as a matter of theory, is held in practice as a truism which it is irrational, if not actually immoral, to question, since it is in the heart of the individual that religion has its throne, and to externalise it in rules and institutions is to tarnish its purity and to degrade its appeal. Naturally, therefore, they formulate the ethical principles of Christianity in terms of a comfortable ambiguity, and rarely indicate with any precision their application to commerce, finance and the ownership of property. Thus the conflict between religion and those natural economic ambitions, which the thought of an earlier age had regarded with suspicion, is suspended by a truce which divides the life of mankind between them. The former takes as its province the individual soul, the latter the intercourse of man with his fellows in the activities of business and the affairs of society. Provided that each keeps to its own territory, peace is assured. They cannot collide, for they can never meet.

History is a stage where forces which are within human control contend and co-operate with forces which are not. The change of opinion described in these pages drew nourishment from both. The storm and fury of the Puritan revolution had been followed by a dazzling outburst of economic enterprise, and the transformation of the material environment prepared an atmosphere in which a judicious moderation seemed the voice at once of the truest wisdom and the sincerest piety. But the inner world was in motion as well as the outer. The march of external progress woke sympathetic echoes in hearts already attuned to applaud its triumph, and there was no consciousness of an acute tension between the claims of religion and the glittering allurements of a commercial civilisation, such as had tormented the age of the Reformation. . . .

* * *

Agreement as to ends implies the acceptance of a standard of values, by which the position to be assigned to different objects may be determined. In a world of limited resources, where nature yields a return only to prolonged and systematic effort, such a standard must obviously take account of economic possibilities. But it cannot itself be merely economic, since the comparative importance of economic and of other interests – the sacrifice, for example, of material goods worth incurring in order to extend leisure, or develop education, or humanise toil – is precisely the point on which it is needed to throw light. It must be based on some conception of the requirements of human nature as a whole, to which the satisfaction of economic needs is evidently vital, but which demands the satisfaction of other needs as well, and which can organise its activities on a rational system only in so far as it has a clear apprehension of their relative significance. 'Whatever the world thinks,' wrote Bishop Berkeley, 'he who hath not much meditated upon God, the human mind and the *summum bonum*,

may possibly make a thriving earthworm, but will most indubitably make a sorry patriot and a sorry statesman.' The philosopher of today, who bids us base our hopes of progress on knowledge inspired by love, does not differ from the Bishop so much, perhaps, as he would wish.

The most obvious facts are the most easily forgotten. Both the existing economic order and too many of the projects advanced for reconstructing it break down through their neglect of the truism that, since even quite common men have souls, no increase in material wealth will compensate them for arrangements which insult their self-respect and impair their freedom. A reasonable estimate of economic organisation must allow for the fact that, unless industry is to be paralysed by recurrent revolts on the part of outraged human nature, it must satisfy criteria which are not purely economic. A reasonable view of its possible modifications must recognise that natural appetites may be purified or restrained, as, in fact, in some considerable measure they already have been, by being submitted to the control of some larger body of interests. The distinction made by the philosophers of classical antiquity between liberal and servile occupations, the medieval insistence that riches exist for man, not man for riches, Ruskin's famous outburst, 'there is no wealth but life', the argument of the Socialist who urges that production should be organised for service, not for profit, are but different attempts to emphasise the instrumental character of economic activities, by reference to an ideal which is held to express the true nature of man.

STRUGGLE AGAINST POWER: NOTES ON INDIAN POLITICAL BEHAVIOUR

Myron Weiner

India's experiment with democratic institutions is still only an experiment. Although India adopted a parliamentary system when she achieved independence in 1947 and has thus far been able to maintain her democratic institutions with much more success than some other newly independent Asian nations, no observer can have complete confidence in their stability. Indians have had to adjust from the role of natives in a country occupied by an imperialist power to that of citizens in a democracy, and the adjustment has not been easy. Traditional attitudes have a way of persisting and are not always readily accommodated to a parliamentary system of government.

This article will examine several pre-independence attitudes that persist in India today as obstacles to the growth of stable democratic institutions and will conclude with a brief discussion of the extent to which some of these attitudes have been weakened. Such traditional attitudes are not the only elements which affect India's democratic prospects and current political behaviour; nevertheless, they are essential factors which are frequently overlooked. They centre on the problems of political power, civil disobedience and the use of violence.

At the core of any democratic system is an acceptance of certain basic attitudes towards law, authority, political power, violence and the like which in the long run may have a greater influence than the formal legal structure in shaping the development of the political system. In the absence of agreement on these attitudes, the legal structure becomes a façade behind which the system really operates; the gap between law and practice grows wide and democracy flourishes in name only. For an example we have only to look at pre-war Japan, where a semblance of parliamentary government existed although government was actually authoritarian.

In attempting, therefore, to discover the prospects for the success of democratic institutions in India, it is important not only to understand the country's rate of economic development, the political strength of the ruling party and the kind of constitutional structure that has evolved; one must also understand those traditional attitudes which affect current political behaviour, and their impact upon the democratic system.

I. POLITICAL CALCULATIONS AND THE RENUNCIATION OF POWER

One prerequisite for a viable democratic system is the existence of a willingness on the part of political groups to make political calculations. That is to say, political parties must be concerned with their electoral prospects and must be willing to make compromises and join or work with other groups in order to increase their own

strength. A viable democratic system has little room for party sectarianism. In India, as in nearly all the other nations of South and South-east Asia which recently won independence, there has been a marked unwillingness on the part of political parties and interest groups to make compromises and merge with one another. As a result, there are in India today some fifty political parties with seats in either Parliament or the state assemblies.

The phenomenon of party splintering is a rather complicated one and cannot be dealt with exhaustively here. Needless to say, it is a phenomenon which may have a profound effect on the growth of stable government. One reason for such splintering is the absence of a basic consensus on the nature of the state. A number of India's fifty parties are Hindu communal parties that reject the secular order. Others are provincial and linguistic parties that have strong allegiances to local caste or language groups. Still others, like the Communist and Marxist left parties, reject the democratic parliamentary framework.

But even among political parties which agree on basic values, political splintering is widespread in India. Underlying this fragmentation is a lack of concern on the part of the political leaders and their followers for the electoral consequences of their actions. When the Socialists broke from the Congress Party in 1948 and when a group of Gandhians broke from it in 1951, the rank and file of both these groups cared little about the effect of their actions on their electoral prospects. Similarly, when the Rashtriya Swayamsevak Sangh (R.S.S.), a Hindu organisation of young people, decided to form a political party, its members were more concerned with maintaining their individuality than with finding some existing party in which they could work effectively. Again, when a few of the leaders of the Hindu Mahasabha and Jan Sangh, the two largest Hindu communal parties, tried to arrange a merger, the rank and file and the other party leaders were more interested in maintaining their own organisations than in increasing their political effectiveness.

An Indian Aneurin Bevan would hardly have fought for long to maintain his position within his party. He would most likely have resigned the moment there was talk of expulsion: he would have condemned the party, its leaders and its programme and proceeded with his own followers to form a new party. He might have given some thought to his electoral chances, but that would have been a secondary consideration. And if he considered his electoral prospects at all, his assessment probably would have been unrealistic.

Why do political consequences or calculations (in the sense of taking electoral prospects into account) play such a small part in the splits that have occurred in Indian political parties? To some extent, this absence of political calculations can be traced to the ambivalent feeling of many Indian politicians and of the party rank and file towards political power as an objective. Perhaps this appears to be a contradiction in terms, since political parties are by nature instruments for achieving power. But it is possible for a political party to exist, to have an ideology, to make demands upon the government, to organise demonstrations and strikes and yet to be unwilling to assume the responsibility of running a government, and to feel at heart that there is something 'dirty' and unpleasant about the whole business of politics.

How widespread such anti-political power feelings are can be easily illustrated. For instance, there are the many members of the Rashtriya Swayamsevak Sangh in the Jan Sangh Party[1] who believe that their objective of changing Indian society and revitalising Hinduism will come about not through the attainment of political power but through propaganda and example. There are sections of the Praja Socialist Party which view the entire political process sceptically. Many of the party members are Gandhians who want to devote their lives to constructive work in the villages or to *bhoodan* work (voluntary land-collecting) and hope thereby to bring about a change in society without governmental intervention. There are Socialist followers of the late M. N. Roy who advocate a society in which there are no political parties and where change would be brought about by activity in villages and factories. And then there are those Socialists who, without being articulate ideological anarchists, nevertheless feel general contempt for the political process and for political power itself. Similarly, anti-power feelings can even be found in the Marxist left parties, many of whose members have abandoned politics to become *sannyasis* (monks) at the famous Aurobindo *ashram* in Pondicherry.

These anti-power feelings, which exist in nearly every opposition party in India,[2] are in large part the product of certain traditional values. Traditionally, high status in India has not been associated with political power. Status was traditionally a function of one's

position in a caste hierarchy in which the Brahman priest rather than the Kshatriya, or ruler-warrior, was at the top. Special status was, however, accorded to those persons, almost irrespective of their caste position, who strove for *moksha* (salvation), an ideal achieved through renunciation, sacrifice and, above all, detachment from this world. The key notion is 'detachment'. One should not have any desire for wealth of power, nor any strong attachments to friends, or one's wife or children. There may be love, but it must be for all creatures or for God, and for no one more than others.[3] One must be detached from all worldly desires.[4]

'Renunciation', 'sacrifice' and 'detachment' are not the ideals by which all men live; only a few can choose this path. For the rest, each man must live in accordance with the duty assigned to him by his caste. In the West, too, the religious life of withdrawal has been confined to a relative few; what makes this notion noteworthy in India is, first, the number of people who leave established positions to devote themselves to the search for salvation; and second, the fact that these ideals are still widely accepted in modern India and have considerable impact even on the lives of those who do not withdraw.

It is also interesting to note that Hindu tradition, while it accords high status to the Kshatriya ruler, holds it important that others should not strive for positions of authority. Each man must accept his own *dharma* (duty) and perform his duty well. Better to perform one's own duty poorly, say the scriptures, than to perform someone else's duty well. Authority is acceptable, but to struggle for a position of authority is not.

As an aside, it might be mentioned here that the Indian tradition of asceticism (which is an extension of the notion of detachment) has posed a dilemma for Gandhian constructive workers who, while they live an ascetic life themselves, advocate the betterment of the villagers' lot. How can one accept the goal of a higher standard of living, but continue to believe in the principles of asceticism? One possible reconciliation – perhaps an unconscious one – seems to be for the Gandhian constructive workers to help the villager without encouraging him to want to improve his lot. Perhaps this is why some of the Gandhian constructive projects have possessed overtones of paternalism in spite of the Gandhian principle of teaching the villagers to help themselves.

The ascetic ideal does, however, have the beneficial effect of discouraging conspicuous displays of wealth. By emphasising asceticism and village constructive work, Gandhi made one of his great contributions to Indian political and social development: he helped to narrow the gap between the educated classes and the villagers, a gap which in other countries has been much wider and has constituted a major obstacle to economic development and social welfare.

Under the British rule, Indian feelings towards government were ambivalent. On the one hand, politics grew in importance as the state took on more functions. But on the other hand, co-operation with the alien British rulers became anathema as the nationalist movement grew. Indians who accepted office were looked upon with some opprobrium. Even after 1936 when the Congress Party agreed to assume office in a number of provinces, virtually the entire leftist section of the party denounced the move. And while one section of the Congress Party talked of the uses to which an Indian national government could be put when independence came, another section under Gandhi continued to emphasise the need to change society by non-governmental activities. With this in mind, Gandhi created a large number of village constructive work organisations. Gandhi's own attitude towards power is illustrated by the story of his reply to a proposal by some of his friends that he become the first Governor-General of India after the achievement of independence. Gandhi is reported to have said that were he to accept office for even one minute, he would negate all of the principles for which he stood.

Today, even with a nationalist government, there are widespread feelings against the 'scramble for ministerial posts', especially among those in opposition. Many Indian intellectuals are prone to feel a certain contempt for those in political office, and to question their motives. These feelings, however, are not so widespread within the Congress Party, since most of those who felt this way, including many Gandhian constructive workers, have left the party. For those who have remained in it, a government post now confers considerable prestige. But this feeling is not generally shared by those outside the Congress Party. Any mention of a 'deal' which would result in the award of ministerial posts is likely to be looked upon not as shrewd politics but as dirty politics. This no doubt motivated part of the criticism of the Socialists – which was voiced by many of the Socialists them-

selves – when the Praja Socialist Party accepted power in the southern state of Travancore-Cochin and formed a minority government with the support of the Congress Party.

These anti-political power feelings not only influence Indian attitudes towards government, but also affect attitudes towards leadership. Politically active Indians, particularly those in the non-Marxist parties, prefer their leaders to be detached even from formal power within their own organisations. The leaders, too, prefer to be outside the formal structure. Gandhi, for example, was not even a dues-paying member of the Congress Party. The leader of the Praja Socialist Party, Jayaprakash Narayan, no longer holds any office in the party. The founder-leader of the Hindu communal party, Ram Rajya Parishad, is Swami Karapatri, a religious leader who is not a member of the party he leads. Golwakar (usually called 'Guruji', or 'Reverend Teacher') is the recognised leader of the R.S.S. and the ideological leader of Jan Sangh, but he holds no position in the latter. Savarkar is the accepted leader of the Hindu Mahasabha but likewise has no official position. All of these men have reputations for 'sacrifice' and 'detachment', and by virtue of their very lack of political attachments are effective in settling the factional disputes that constantly plague every Indian party.

II. STRUGGLE FOR POWER

In a society whose members have been wielding political power for only a short time, one can expect few rules about the just and unjust ways of using power. In India, this situation is compounded by the absence in tradition of any notion of a universal system of ethics. There are no Ten Commandments: each man has his own *dharma*, and if he wants to achieve a better position in his next life, he must act in accordance with his prescribed role. Each *dharma* prescribes its own ethical rules. While on the one hand, for example, there is a tradition of non-violence among some sects of Hinduism, on the other hand, in the *Bhagavad Gita*, one of the most famous and popular religious texts, Krishna tells the warrior Arjuna that he must kill because it is his duty to do so. Another famous work, the *Arthasastra*, describes the ways of winning, keeping and using power in a manner which outdoes Machiavelli's *Prince* in its apparent disregard for moral principles. But these seemingly non-religious attitudes are none the less part of the religious tradition. The ethical rules for the behaviour of a Kshatriya ruler are different from those of a Brahman priest, just as their *dharma* differs.

When power does become a goal in a society in which an ideal is the renunciation of power, and where no system of ethical rules has developed concerning its uses, it is frequently sought in chaotic fashion and used in 'unethical' ways (in the conventional Western sense of the word). This may help to explain why in the 1952 elections nearly six thousand independents contested seats for the various state legislatures – averaging nearly two independent candidates per seat – and why the Congress Party was flooded with office-seekers shortly after independence. This may also help to explain why before the 1952 elections state governments were frequently handicapped by politicians who used their government posts to make money and aid relatives and friends, and why corruption and nepotism have often been practised openly and in rather crude ways.

The struggle *against* power therefore has its converse in the struggle *for* power. Each reinforces the other. The views of the anti-power personality, especially, are reinforced by observation of the behaviour of those who struggle for power. Lack of familiarity with democratic rules breeds irresponsibility and, in turn, irresponsibility breeds a greater contempt for the democratic system.

III. LAW AND CIVIL DISOBEDIENCE

In pre-independence India, civil disobedience, or what Gandhi called *satyagraha* (truth force), was a major weapon against the British rulers. So long as this refusal to obey government law was directed against an alien ruler, law-breaking was accepted as a weapon by virtually all politically active Indians. But the British emphasis on the rule of law had a profound effect on Indian intellectuals, especially those who entered the legal profession. Gandhi himself was a lawyer and although he urged civil disobedience against the British, he emphasised respect for the law. Civil disobedience was frequently, although not always, directed against administrative rulings rather than the law, and it was often paralleled by legal action in the courts by Indian barristers. But

whatever Gandhi's intentions, the overall effect of *satyagraha* could only be a receptivity to the violation of law on the part of the Indian masses. So long as law was the product of an alien rule, British attempts at making law legitimate through reference to democratic notions could hardly be convincing to most Indians.

But with the achievement of freedom, the use of *satyagraha* came into question. Does organised law-breaking, even when it is non-violent, have a place in a democratic society? The Socialists, who during the past few years have accepted many of Gandhi's principles, conducted *satyagrahas* on several occasions in Uttar Pradesh, Bombay, and elsewhere against what they considered to be injustices on the part of the government. Eviction of tenants and high taxes – especially high land or irrigation taxes – have frequently been the targets of Socialist civil disobedience movements. While Gandhians have criticised the misuse of *satyagraha*, in principle they continue to support it. *Harijan*, the official Gandhian weekly, published a note by its late editor, K. G. Mashruwala, in defence of *satyagraha*:

> Satyagraha – insistence on truth – does not depend upon the form of Government. . . . Even if it needs to take the form of individual or organised non-violent civil resistance of an unjust law or order, or of non-co-operation with a general or particular evil, or of a resort to fast, a *satyagrahi* [one who takes part in a *satyagraha*] cannot avoid it even in a democracy . . . it has to be taken after attempts of bringing about conversion through arguments and discussion and other legal remedies have failed . . . in a well-governed democracy, there should be few occasions for resorting to it. But it cannot be entirely ruled out, in a proper case.

The same issue of *Harijan* published a reply by Pandit Pant, then Chief Minister of Uttar Pradesh and now Home Minister in Nehru's cabinet:

> in matters of conscience – on spiritual issues – one is always entitled to offer Satyagraha. But, for a democratic set-up, I had always stressed the importance of settling political differences by argument and persuasion. . . . If Satyagraha is to be resorted to or threatened every time there is a difference of opinion, then public opinion does not receive the respect that is due to it. . . . It would, no doubt, be a matter of conscience if one were forced to offer Satyagraha in a case where the issues are clear, and the Government's stand is patently untenable, and where before any recourse to Satyagraha the pros and cons have been examined in an impersonal, detached and dispassionate manner, and all other remedies have been exhausted. He who decides on Satyagraha should have a clear conscience and an open mind: his judgement should be detached and his decision should be unbiased. But in the midst of political partisanship, how many are there who can, or will, act in this manner? Democracy would lose much of its significance if, in the name of Satyagraha, everyone were to take it as his unfettered right to incite the masses, as much as possible, against the system of Government established by the people's elected representatives in a democratic State, in order to belittle and resist the law of the land in a concerted manner. . . .

In spite of the protests of the Indian government and of most Congress Party leaders, *satyagraha* has been applied on a number of occasions by several political parties since independence. The *satyagraha* in Andhra, culminating in the fast-unto-death by a Telugu leader, was a major factor in the government's decision to sanction the formation of the state of Andhra out of the Telugu-speaking areas of Madras.

Non-violent law-breaking likewise played a large part in the Jammu and Kashmir agitation in early 1953. The major Hindu communal organisations, the Jamma Praja Parishad, Bharatiya Jan Sangh, the Hindu Mahasabha and Ram Rajya Parishad, joined together to agitate for 'full and unqualified' accession of Jammu and Kashmir to India. Jan Sangh and the Hindu Mahasabha were particularly active in Delhi, where they organised demonstrations in defiance of police orders. Syama Prasad Mookerjee, the Jan Sangh leader, defied the government by entering Kashmir without the necessary government permit. He was subsequently arrested, as were a large number of leaders and party members who took part in the *satyagraha*.

Civil disobedience movements have occurred throughout India since independence, and the Indian government has in virtually every instance used the power of the State to suppress them. In spite of the fact that India is no longer subject to foreign

rule, most organisations outside the government or the Congress Party still believe that *satyagraha* is a legitimate technique for achieving one's objectives. Their feeling is that when they fail to achieve their objectives through regular channels – petitions, elections, delegations to the government and the other devices for exerting influence within a democratic society – non-legal devices ought to be employed. Those who think this way do not accept the notion of the sanctity of law which underlies the power of most Western democratic governments, where legal rules legitimise the authority of government.

In India, political power has not yet been made legitimate. The justification for power by reference to *dharma*, whatever role it once may have played, ceased to be a justification with the arrival of the British. Instead, the British introduced the concept of 'justice under law'. Although this idea made considerable inroads on Indian thinking during the nineteenth century, so long as the national government was under foreign rule, the bulk of the nationalist movement – as we have already noted – saw no 'justice under law'. Thus at the time of independence, India was not, to use Max Weber's term, a society based on 'rational–legal authority', that is, a society in which authority is made legitimate by reference to law. Opposition party members and apparently large numbers of peasants and workers do not yet feel the restraints of law. Injustice, whether it be by act of government or of landlords, is greeted by a readiness to violate law or contract.[5]

IV. LEARNING THE RULES

It is not surprising to find that Asian countries, having suddenly adopted political structures of a Western type, are struggling to accommodate their traditions to the new institutions. As this article has indicated, despite India's success thus far with democratic institutions, there still persist traditional attitudes which do not always facilitate adjustments. India's problem is the same as that of many other Asian countries: how to adapt traditional attitudes to a rational–legal democratic political system.

There are, however, a number of elements at work in India that tend to ease her adjustment to democracy and make the prospects for her political future more hopeful than those of many other Asian countries.

First of all, the intensity and duration of the Western impact have been greater in India than in nearly any other Asian country. Traditional Indian attitudes and institutions have been profoundly affected. A large, Westernised university system has been established, and English has become the language of the educated community. The administrative and military services, among the most efficient in Asia, are Westernised. The number of lawyers inculcated with Western legal notions is very large, probably considerably larger than one finds, for instance, in Japan. Furthermore, even before 1947, many Indians had gained considerable experience in the State legislatures and in high government posts.

It is not possible here to analyse exhaustively the extensive impact which the West has had on India and the remarkable modifications which have been made not only in social and political practices, but in religious and political theory as well. What is important to note, however, is that Hinduism has shown a remarkable capacity to assimilate and integrate Western notions. In large part, traditions are being modified rather than supplanted.

Secondly, those traditional attitudes which are not consistent with democratic rules are beginning to break down as Indian politicians and the party rank and file grow accustomed to a parliamentary system. It takes time to learn new rules, and only since the 1952 elections have most Indians begun to feel that they are living under a new political system. For example, minor parties are increasingly recognising the necessity to make compromises and merge with other parties to augment their strength.

In the 1952 elections, the splintering of the opposition allowed the Congress Party to win a clear majority of the seats in Parliament with only 45 per cent of the popular vote. But the absence of proportional representation, and the presence of a system in which a plurality vote means victory, may eventually force some of the smaller groups either to merge or to resort to extra-legal, extra-parliamentary devices to avoid extinction. In the long run, the successful operation of the parliamentary system is likely to bring about a change in Indian attitudes towards political power, especially as government successfully meets the needs of the community. Likewise, if the gap between expectations and fulfilment is narrowed, and if discontented groups discover that they can achieve results through the regular channels of a parliamentary system, there is likely to be a decrease in civil disobedi-

ence. Even now, although civil disobedience is practised, there is considerable criticism of it and those who have violated the law have found it necessary to explain publicly why in a parliamentary system they must resort to civil disobedience.

The essential point is that the sheer survival and operation of stable democratic government in India are in themselves perhaps the major factors that will contribute towards the breakdown of traditional obstacles to the democratic system and towards persuading opposition groups to accept the system. Political stability is a function of many complex factors, but underlying them all is the simple matter of survival for a substantial period of time. 'Nothing succeeds like success' still makes a good political axiom. If the present democratic institutions of India continue to function effectively, if law and order are maintained and stable government continues in the various states as well as in the central government, faith in the system is likely to grow. If Indians come to recognise that their aspirations can be fulfilled within the democratic system, then they are likely to operate within the rules.

Notes

1. Jan Sangh is the largest Hindu communal party and the fourth largest party in India. The other three are the Congress Party, the Praja Socialist Party and the Communist Party of India.

2. The one exception being the Communist Party, perhaps the most Westernised of all the Indian parties and one which is completely concerned with the objective of winning power.

3. See, for example, Gandhi's autobiography, in which Gandhi says, 'He who would be friends with God must remain alone, or make the whole world his friends.'

4. A disciple of Gandhi wrote that a central teaching of the *Bhagavad Gita* which played a major part in Gandhi's thinking was the ideal of *nishkamakarma* (action without the desire for result).

5. Although Gandhi personally rejected the use of government to achieve the kind of society he envisaged, he did contribute a device for channelling resentment against government. While he was apparently prepared, *under any kind of government* (even in a democracy), to violate laws he considered unjust, violations were to be accomplished in a non-violent way. In a society such as India, where there are few self-restraints to prevent violation of law and where the population is becoming increasingly aware of the existence of 'injustice', the choice is not merely between abiding by or violating law, but is frequently between non-violent and violent law-breaking. Gandhi's contribution was a special set of non-violent, albeit non-democratic, rules by which laws could be violated. The degree of violence might be far greater in India today had not Gandhi provided such a set of rules. As Lois Murphy noted in a recent study, in a society in which there is a tradition of little violence and where consequently there are no rules that put some restraint on its employment (rules such as 'Fight like a man', 'Don't hit below the belt', and other canons of 'fair play' and 'sportsmanship'), when violence is committed, it may be even more extreme and cruel than in a society where its use is customary.

WORK LIFE AND POLITICAL ATTITUDES: A STUDY OF MANUAL WORKERS

Lewis Lipsitz

FOR centuries men have speculated about the human consequences of work. Slowly, a considerable body of information has begun to accumulate concerning the relationships between people's jobs and other aspects of their lives. Investigators have pointed out the connections between certain types of jobs and certain personality disorders, job satisfaction, productivity, attitudes towards union and management, social relations on the job, leisure activities, and other things. Extending such findings, this study concludes that a particular job situation can have important effects on a man's political outlook.

Political studies have long classified individuals according to occupation, yet there have been extremely few efforts to penetrate within specific occupational categories to discover the causal triggers of occupational attitude differences. Political scientists have, on the whole, been content to work with relatively large groupings such as 'semi-skilled' manual workers, 'unskilled' manual workers, etc. Several efforts have been made, however, in the direction this study takes: exploring specific job characteristics and their effects on political attitudes. Seymour Lipset cites a work that established a relation between size of factory and voting Communist or Nazi in pre-Hitler Germany. William Kornhauser observes the connection between the rate of expansion of a particular industry in the early stages of industrialisation and the political ideology of the emerging labour movement. In such findings, we move somewhat closer to establishing a causal relationship between something about a particular work situation and certain attitudes or behaviour of the job holders.

But more specifically, concerning the problem of job satisfaction, Lipset, writing in 1960, observed that it was yet to be shown that 'job satisfaction and creativity contribute independently to political behaviour over and beyond differences in status and economic conditions . . .'. The present study indicates that a particular work situation, that of the automobile assembly-line, affects the political and social attitudes of the automobile workers involved. The ways in which the concrete work situation affects the worker's political attitudes are determined by the technology and social setting of the job itself. Specifically, assembly-line workers are found to be more fatalistic, more punitive, and more politically radical than other workers of comparable salary and education who work in the same plant.

I

Before proceeding to the findings themselves, it may be helpful first to consider briefly the nature of assembly-line work, in comparison with other occupational types.

The social and psychological consequences of work have been much debated. The factory is a symbol of modern life, and the assembly-line in particular has often been thought to embody in extreme form the worst tendencies of industrial work. Other analysts have argued, to the contrary, that modern industrial societies offer more individuals more chances for genuine work satisfaction than any previous societies. It would serve little purpose here to get involved in this controversy, for even if the issues raised remain unresolved, certain facts about modern work-life relevant to our inquiry are quite clear. First, manual workers generally indicate lower work satisfaction than non-manual workers. Second, within the manual group, satisfaction varies with skill: the higher the skill, other things being equal, the greater the satisfaction. Table 27.1 below provides some frame of reference for dealing with the question of job satisfaction in manual and non-manual jobs.

Beyond the question of how workers respond when asked if they are satisfied with their jobs, a more important question concerns the *effects* of work – what impact the job has on a man's life outside of the factory. Many writers and analysts have speculated about this. Much of the analysis of industrial work emphasises its destructive or negative aspects and focuses attention on the elimination of skill and the increase of simple, repetitive, assembly-line style work. Yet there have been only a few studies of assembly-line jobs.

TABLE 27.1

PROPORTION IN VARIOUS OCCUPATIONS WHO WOULD CHOOSE SAME KIND OF WORK IF BEGINNING CAREER AGAIN[1]

Professional occupations	%	*Working-class occupations*	
Mathematicians	91	Skilled printers	52
Physicists	89	Paper workers	52
Biologists	89	Skilled automobile workers	41
Chemists	86	Skilled steelworkers	41
Lawyers	83	Textile workers	31
Journalists	82	Unskilled steelworkers	21
		Unskilled automobile workers	16

Source: Robert Blauner, 'Work Satisfaction and Industrial Trends in Modern Society', in *Labour and Trade Unionism: An Interdisciplinary Reader*, ed. W. Galenson and S. M. Lipset (New York, 1960) p. 343.

The development of the assembly-line was part of what one observer has called the 'second industrial revolution'. Although overhead conveyors were employed from the 1870s on in the American packing-house industry, they were not introduced on a large scale until 1914, when Ford's Highland Park, Michigan, plant became conveyor-equipped. After 1920, assembly-lines were rapidly adopted in many areas of heavy industry.

Charles Walker and Robert Guest, who made probably the most thorough study of assembly-line work, characterised the 'pure' assembly-line job as involving the following six elements:

1. Mechanical pacing of work.
2. Repetitiveness.
3. Minimum skill requirement.
4. No choice in the use of tools or techniques.
5. Minute subdivision of the product worked on.
6. The need for constant surface attention.

The expectation that men would dislike assembly-line work is borne out by current research. Consistently, investigators have found that conveyor-paced jobs in mass-production industries show lower levels of satisfaction than other manual and non-manual jobs. Blauner concludes that assembly-line work is more disliked than any other major type of work. Within the automobile industry, which is the setting of this study, the position of

the semi-skilled line worker is difficult. Possibilities of advancement within the ranks are extremely slim. Wage differentials between the highest- and lowest-paid production jobs are small.

Automobile assembly-line workers have little control over the pacing of work. The average worker on the line performs no more than five operations, remains at his station at all times and repeats the same operations about forty times an hour. For most workers, the work cycle is between one and two minutes, regulated by the conveyor speed.

Characteristically, the assembly-line worker is alone on his job. The technology of the line increases the worker's sense of impersonality and presents barriers to group cohesiveness. Moreover, there appears to be no adjustment to the assembly-line as time passes. The longer a man has held an assembly-line job, the more likely he is to have an unfavourable attitude towards all aspects of work experience.

What do we know about the effects of assembly-line work? Compared to other groups of manual workers, assembly-line workers show few grievances, little use of pressure tactics and much internal disunity and evidence of suppressed discontent. They permit leadership to gravitate to aggressive individuals with a strong need to dominate. They accept certain elements of their work life as inevitable, and they lack self-confidence.

When Walker and Guest compared workers holding 'pure' assembly-line jobs with another group of workers whose jobs were more skilled, varied and autonomous, they found that absenteeism was much higher on repetitive, conveyor-paced jobs; repetitiveness was negatively correlated with interest in work; limitations on talking were a source of frustration to line workers; mass-production methods tend to increase the worker's sense of anonymity. My study indicates that the differences Sayles and Walker and Guest find between different groups of semi-skilled factory workers carry over into the area of political attitudes.

II

Three samples of workers were used here as the basis for comparisons:

1. Assembly-line workers, or men with jobs approximating those of the assembly-line.
2. Repair, relief and utility workers.
3. Skilled maintenance workers.

Repair, relief and utility workers were chosen as the semi-skilled group for purposes of comparison because their jobs provided them with greater autonomy and variety than the line workers had. Repair jobs, for example, could not be time-studied and were not completely repetitive. Individual judgement and some choice of tools were required. Relief and utility workers both performed as stand-ins for men along the assembly-line, but were competent to handle many kinds of line jobs within a given department of the shop. Their work was more varied and likely to lead to a stronger sense of pride and a deeper interest in the task.

The repair, relief and utility workers, like the assembly-line operatives, would normally be classified as semi-skilled in a social class or occupational breakdown. Skilled workers were also included since they would furnish an important comparison with both semi-skilled groups and might exhibit in greater detail or more extensive development tendencies found in the other groups.

Names were selected at random from the files of the local union. Men willing to participate were interviewed in their homes. The interview questions were almost entirely open-ended. In addition to the interviews, each man completed a questionnaire which incorporated various attitude scales and other questions. All of the men interviewed were married, native-born and white. The three groups were closely comparable in many respects. Educationally, all three groups averaged between nine and ten years of education completed. A majority of each group was Catholic. The mean age of each sample was between forty and fifty. Assembly-line workers interviewed had been on their present jobs an average of six years; repair workers, twelve years; and skilled workers, thirteen years. Pay rates in the two semi-skilled groups varied between $2.40 and $2.74 an hour, while in the skilled group the rates ranged from $2.96 to $3.39.

The assembly plant in which these men worked was located in Linden, New Jersey. At the time of these interviews, between 2900 and 3000 men were employed at the plant. The general effort was to study intensively the political orientations of a relatively small group of men, to gain some insights of general applicability that could then be tested on a larger sample, more systematically.

III

The three work groups showed attitude differences in five areas: the job; the union; a fatalistic view of social life; the extent of preoccupations with economic problems; and the extent of sympathy and tolerance.

The job

Assembly-line workers view their jobs with a mixture of anger and resignation. Almost all of them complain bitterly of overwork, physical strain and monotony. They feel subjected to undue pressures because of the excessive speed of the line, but they feel powerless to control this speed. Relationships with supervisors are marked by inequality and anxiety. These men feel expendable. They know there is nothing unique about their performances at work.

All repair, relief and utility workers have ambivalent attitudes towards their jobs. Many complain of pressure and overwork, of physical and emotional strain. Yet they also note compensating factors: variety, skill and the very important fact that the men on the assembly-line have much more unpleasant jobs than theirs.

The fifteen skilled workers present a picture of their jobs in extremely vivid contrast to those offered by the other two groups. Not a single man dislikes his job, and the degree of satisfaction is high in most cases. By and large, these men note as important the fact that they set their own pace of work; they are not closely supervised. They speak of pleasure in the creative use of their skills. They feel themselves in a position to perform useful and appreciated services for others. They have relatively good relations with supervisors. They show no sense of resignation to the inevitable; they know they can leave the corporation since they possess skills needed elsewhere. They have a certain self-confidence built on the knowledge that they have something valuable to sell.

The following tables will convey some notion of the differences in job attitudes among the three groups.

The union

The twelve line workers show slightly favourable attitudes towards unions. Only four are altogether favourable. Five others are ambivalent, citing both good and bad features of unionism. Finally, three men are largely hostile towards unions.

Those with anything good to say about unions note 'protection' as the most important union function; indeed, it is the one positive union function the line workers cite. They see the union as the only force that blunts the exploiting bent of the company. Several of the men mention seniority as an especially important kind of protection. Those who criticise the union concentrate their attention on the idea that unions are not controlled by the workers but rather by a clique which serves its own interests. They assert unions are corrupt and uninterested in the rank-and-file. As a

TABLE 27.2

MAJOR COMPLAINTS ABOUT JOB

	Per cent citing					
Occupation	*Pressure or over-work*	*Physical or mental strain*	*No skill or mono-tony*	*Bad super-vision*	*Exploi-tation*	*N*
Line	75	50	75	58	50	12
Repair	50	43	7	36	43	14
Skilled	13	13	0	20	7	15

TABLE 27.3

MAJOR SOURCES OF JOB SATISFACTION

	Per cent citing					
Occupation	*Variety or skill*	*Good super-vision*	*Auton-omy on job*	*Better than line*	*Any kind of satis-faction*	*N*
Line	0	0	0	0	42	12
Repair	78	28	0	57	78	14
Skilled	80	67	27	67	100	15

TABLE 27.4

ATTITUDE TOWARDS LEAVING PRESENT JOB

Occupation	*% who would leave if possible*	*% who feel stuck*	*N*
Line	91	75	12
Repair	71	64	14
Skilled	27	13	15

group, these men are not closely attached to their union. They give no indication of a sense of participation or solidarity.

The views of the repair, relief and utility workers resemble those of the line workers in several respects, but differ in several others. These workers show more favourable feelings towards unions generally and also towards their own union. Eight of the fourteen men have pro-union attitudes, while six are ambivalent. None is completely anti-union. Like the line workers, they stress

TABLE 27.5

OVERALL ATTITUDE TOWARDS UNIONS

	% *generally favourable*	% *pro-U.A.W. critical of others*	% *ambivalent*	% *generally unfavourable*	*N*
Line	33	0	42	25	12
Repair	28·5	28·5	43	0	14
Skilled	46	0	45	0	15

TABLE 27.6

ATTITUDES TOWARDS WALTER REUTHER

	% *favourable or ambivalent*	% *unfavourable*	*N*
Line	25	75	12
Repair	36	64	14
Skilled	86	14	15

the importance of 'protection' as a union function, but four of the men also note union benefits such as vacations, company-paid insurance, etc. Within the repair group, only one man accuses unions of being oligarchical. Criticisms move in other directions and seem to be less intensely felt than among the line workers.

Much as might be expected, the skilled maintenance men also show relatively favourable attitudes towards unions. Seven are wholly pro-union, while the eight others are ambivalent. This group, too, emphasises the 'protection' the union offers, but stresses union benefits even more than the other groups. None of these skilled workers show the sense of abandonment found in the line sample. And none of them criticise unions as oligarchies.

Attitudes towards Walter Reuther, President of the United Auto Workers, are quite revealing. Both the line and repair workers are sharply critical of their union head. A majority in both groups condemns Reuther as a man out for himself, personally ambitious and uninterested in the fate of the ordinary worker. Even men who praise the U.A.W. as a whole, and unions generally, are wholly negative where Reuther is concerned. The skilled workers, by contrast, are overwhelmingly favourable towards Reuther, praise his skill and intelligence, as well as his honesty and dedication. In all, attitudes towards Reuther are more extreme (both positive and negative) than attitudes towards unions generally, or the U.A.W. in particular. The line workers, who are most unhappy at work, are least happy with unions and with Reuther. The skilled workers, who are most satisfied on the job, are largely pleased with unions and with Reuther. The repair group occupies a middle position in both areas.

Fatalistic attitudes

It is sometimes argued that fatalistic attitudes do not characterise industrial man. Fatalism, it is held, is characteristic of the peasantry, lending support to rigid class distinctions and to the idea that the political order is not to be tampered with.

No doubt citizens of industrial societies are not as fatalistic as their ancestors. Yet it is also clear that the social and technical structures of industrial societies themselves produce certain fatalistic orientations, and that these are related to the specific kinds of work men are engaged in. Those workers who are conscious of having the least control over their own lives at work show the most pronounced tendency to view the social and political worlds as unalterable.

To say the very least, it is difficult to determine the degree of fatalism comprised in an individual's outlook. In an effort to find an approximate means of measurement, individuals were scored on a fatalism scale composed of four open-ended interview questions. Table 27.7 below indicates the results of this scoring.

There are clear differences among the three work groups and these differences are in the expected direction. The assembly-line workers are the most strongly fatalistic. Though there is not space here to explore the types and nuances of fatalistic attitudes, a few important points need to be made. First, as a group, these men are most strongly fatalistic about the problem of war. The idea of the inevitability of aggression fits well with the picture these men

TABLE 27.7

FATALISM SCORES BY JOB GROUP*

Occupation	*Number of fatalistic answers* None	One	Two	Three	Four	*Mean score*
Line	0	2	3	5	2	2·58
Repair	3	2	5	2	2	1·86
Skilled	3	6	3	3	0	1·40

*The four questions that made up the scale were:

1. Will men and nations always fight wars with one another?
2. Will there always be poverty in the world?
3. Do you think the ordinary man is helpless to change some aspect of government he doesn't like, or is there something he can do about it?
4. Some people say they can plan ahead for long-range goals and then carry out their plans. Others say, 'Whatever's going to be is going to be and there's no sense planning.' How do you feel about it?

The four questions form a Guttmann-type scale with a reproducibility coefficient of 0·91, and a minimal marginal reproducibility of 0·671.
Analysis of variance significant at 0·05.
t test for line and repair significant at 0·15.
t test for line and skilled significant at 0·01.

have of human nature. Second, fatalistic attitudes are weakest in the area of the political potency of the ordinary citizen. These men retain the notion that a letter to one's Congressman will change things. They hold to this idea despite the fact that on other questions they indicate their belief that public officials don't really care about the common person.[2]

The non-fatalistic workers in the sample stand out most strongly from their fatalistic fellows in two respects. First, they are indignant rather than merely sullen and resigned about conditions they find unacceptable. Second, the non-fatalists tend to blame themselves as well as others for social failures. They emphasise the possible triumph of justice, even if only in the distant future, and the possibility of humane and rational behaviour, even though many of them consider such behaviour unlikely.

Economic preoccupations and political radicalism

The assembly-line group is more politically radical than the other two samples of workers. This manifests itself in favourable attitudes towards governmental economic controls and antagonisms towards private ownership and the privileged orders; it is closely related to a strong tendency among the line workers to emphasise economic problems and interpretations.

The distribution of attitudes along the liberal–conservative continuum is in the expected direction. The assembly-line workers are the most radical, the skilled workers the least. The patterns for the three groups bears a close resemblance to the scores for skilled, semi-skilled and manual workers which Centers found in a nation-wide sample in the late 1940s.

TABLE 27.8

WORK GROUP SCORES ON LIBERALISM–CONSERVATISM SCALE[3]

	% *ultra-conserva-tive*	% *conserva-tive*	% *indeter-minate*	% *radical*	% *ultra-radical*	*N*
Line	0	8	33	42	17	12
Repair	8	15	54	23	0	13
Skilled	13	33	27	20	7	15

$X^2 = 0{\cdot}10$.

Symptomatic of their economic preoccupations, two contrasts emerge among the work groups in which the line workers attach a greater emphasis to economic considerations. First, they tend more strongly to view economic problems as the most important ones confronting America today. Fifty-eight per cent of the line group ranks economic problems most important as compared with 36 per cent of the repair group and 33 per cent of the skilled workers.

The kinds of economic problems mentioned by the different groups were also quite distinct. All of the men in the assembly-line group who mentioned economic problems at all concerned themselves either with unemployment or poor working conditions, both problems very close to home. In contrast, men in the other two groups raised questions about high taxes, foreign aid, welfare excesses, salary differentials and the demise of free enterprise.

Second, the concern with economic difficulties shows itself in responses to a question that asks about the nature of the 'chains' men feel. The question was phrased, 'A man once said: "Man is born free, but everywhere he is in chains." Does this mean anything to you?' Table 27.9 arranges the responses according

to whether the individual saw Rousseau's 'chains' in economic terms.

Table 27.9 is significant not only because the line and repair workers tend to perceive 'chains' in economic terms, but also because they tend to see these 'chains' as oppressive. Remisik, a

TABLE 27.9

RESPONSES TO ROUSSEAU

	% *offering economic interpretation*	% *offering non-economic interpretation*	*N**
Line	86	14	7
Repair	77	23	13
Skilled	27	73	11

*Percentages are based only on those who offered some interpretation and therefore exclude the 'don't knows'.
$X^2 - 0{\cdot}02$.

repairman, typifies much of the commentary: '. . . has to earn his bread. He's chained to a job of some sort, regardless of whether it's here or any other concern. . . .' The 'chains' these men see range over bills, the regimentation of factory life, and the income tax. On the other hand, most of those who offer non-economic interpretations think of 'chains' without the element of oppression. Harris, a maintenance painter, serves as a good example: 'Yeah, everybody is born free, true – but then we acquire chains. It has to be. There's law, regulations that we have to abide by. . . . We're always in chains, but chains can get awfully heavy in some of these countries.'

Sympathy and tolerance

None of the questions in the interview schedule was designed directly to explore dimensions of sympathy or tolerance. Yet several of the questions seemed to elicit punitive responses from some of the men but not from others. All of these questions touched on attitudes towards 'out-groups'. The four questions were:

1. Why are people poor?
2. What should our policy be towards Russia?
3. Do you think it was a good idea to drop atomic bombs on Japan in World War II?
4. How do you feel about capital punishment?

Though the evidence is very skimpy and impressionistic, it appears that the assembly-line group is less sympathetic and tolerant than the others. In all of the four cases, the line workers are slightly more punitive as a group. They tend more strongly to blame the poor themselves for poverty. They were more likely to approve the bombing of Japan without showing any sympathy for the Japanese, and to recommend a policy of toughness towards Russia. Finally, they were somewhat more likely to favour the death penalty.

These findings are worth noting largely because they tend to confirm previous conclusions about authoritarianism and social class. Within this group of workers, punitive attitudes are associated with work frustration and lack of control over work life. These facts point up the possible significance of everyday frustrations on conditioning attitudes towards out-groups.[4] The possibility that such frustrations might play an important role in shaping attitudes is one that psycho-analytic research has left open for the moment.

IV

In two important areas the three groups show attitudes that are much alike but are important to describe none the less in order to give greater definition to the picture of working-class views drawn thus far. First, a majority of the men in all three groups emphasise the manipulative nature of much of social life. They express concern about the exploiting and unfair practices of their corporation and of big business generally. Moreover, approximately half the men express concern about two-faced, deliberately deceptive tactics of the government and mass media which they see as closely related. Table 27.10 shows the distribution of such attitudes in the three work groups.

Where big business is concerned, these workers see manipulative practices everywhere. A few believe big business or the rich run the United States. Others are not quite so explicit, but believe that big business is free of the restraints that bind ordinary mortals: businesses can initiate wars, bribe judges and representatives and manage to get the government to follow their dictates. In some of these men, hostile attitudes towards big business, and especially towards their own corporation, attain a frenzied pitch and a deep bitterness.

Most of the men feel that the corporation acts as it does because its primary concern is profit-making. Some see businesses caught in a competitive rat-race in which it is do or die. The methods of the rich, as these workers describe them, are not subtle. These men simply affirm the idea that 'money talks'. Very few of them have any praise for big business. Most simply accept it as a fact of life, unable, like the line worker Tencio, to visualise a world without it: 'Well, without big business, what could the working man do? He gotta work for big business.'

TABLE 27.10

PER CENT PERCEIVING MANIPULATIVE PRACTICES BY BIG BUSINESS OR CORPORATION X, AND BY GOVERNMENT OR MASS MEDIA *

	See manip. by Corp. X or big business (%)	*See manip. by govt. or mass media* (%)	*N*
Line	92	50	12
Repair	79	43	14
Skilled	60	47	15

*The specific questions asked were:

1. How do you feel about big business?
2. How do you feel about Corporation X?
3. Do you feel there's much corruption in politics? Why?

It is not particularly surprising to learn that many of these working-class men are critical of big business. But it is very much another thing to find that almost half of the 41 men are critical of manipulative and deceptive aspects of American government and the mass media. The chief theme of the critics is that the government is immune to popular pressures. The general picture is of a group of men who arrange to keep themselves in power and who don't care about the requests of the common people. The means of governmental manipulation are varied: nepotism, misleading propaganda, suppression of information, bribery, silencing of dissent. The credibility of the media is to some degree tied up with the credibility of the government. Several of the men feel that the radio, papers, television only convey what they are told to. They are tools, willing or unwilling, of the politicians. Five men criticise the media as manipulative in their own right. All who discuss the media do so in terms of their deceptive political content.

Some of the workers who are concerned about manipulations tend in the direction of a conspiracy theory of social control. Yet few of them actually express anything resembling a theory of a hidden oligarchy or oligarchies. At most, seven men give any clear indications of holding to such a picture of American society. In all seven cases, it is the 'rich' or 'big business' which is said to be the ruling elite.

The remaining two-thirds of the men who criticise various manipulative practices give no indication of adhering to a power-elite theory. They criticise business, government and media for attempting to manipulate the common man, but their criticism is specific and piecemeal. They do not have any clearly formulated picture of social life, nor any precise notions about who is to blame for the seamy side of things. They are critical of the excessive power of big business and yet do not believe that big business runs the country. They can believe that people in the government attempt to mislead the public without going sour on government as a whole.

Yet behind their vagueness and their sense of their own ignorance, most of these men actually have developed an explanation for the practices they criticise. First, they see a society directed towards profit-making. Second, they are aware of wide disparities of power. Out of these elements, these men have constructed a suspicion of human nature. It is human nature and nothing less, they say, which leads to the malpractices of business and government. Thus, it is not a particular group of evil men who are to blame, but rather the selfishness supposedly inherent in almost all of us.[5]

These workers generally accept the profit-motive as a satisfactory explanation of business behaviour. But this does not account for governmental behaviour. Why these failures, large and small, in the democratic system?

A substantial majority of the men believe that politics, like most other aspects of our society, is basically a business. Three-quarters of the line workers, 57 per cent of the repair sample and 60 per cent of the maintenance men see politicians as predominantly self-seeking. They see elections and platforms designed to fool the public; politicians seeking to perpetuate their own

domination; stealing, nepotism, the wasting of public funds, etc. A government composed of such individuals could not possibly be very attentive to the kinds of problems these workers are concerned about. It is not surprising then that they see failures in the political system. The selfish nature of politicians springs from their unfortunate and seemingly indelible humanity.

The second important area of similarity among the three groups is political participation. Unexpectedly, all three groups were strikingly alike in regard to voting, discussion of politics, and interest in political affairs. A majority of the men in all three groups reported talking politics at least sometimes. Only one man indicated he was not interested in presidential elections. Only a handful felt that the outcome of the 1960 presidential election made no difference. All of the men said they had voted in 1960.

Only in two areas did any tentative difference appear among the three work groups. First, the skilled workers were more likely than the other groups to participate in political activities beyond voting. Two-thirds of the skilled group had contributed money, gone to party meetings or actively campaigned in recent years. In the repair group, the figure was 43 per cent, and in the line group, 36 per cent.

Second, the quality of partisan commitment varies considerably among these groups. Very crudely, there are two sorts of commitment: positive and negative – being in favour of, or being against. Most notably, the line workers phrase their political commitments in terms of their aversion for the Republican Party as well as their preference for the Democrats. Five of the twelve line workers speak of their ill-feeling towards the G.O.P. This ill-feeling runs the gamut from quiet discontent to raw bitterness. Dolski, a line worker, speaks with particular resentment on the subject of poverty: 'Why do we have poor people in this state? I'll tell you: Republicans got in there. They kept everything for themselves. Now we got the poor in. . . .'

V

Much of the evidence set out in the preceding sections is impressionistic and statistically insignificant. Yet the weight of the evidence points in a single direction.

It is clear that the two groups of semi-skilled auto workers hold different views in various areas of life. First, the men who work on repetitive, assembly-line jobs dislike their work more and are more hostile towards and alienated from the union. Second, the assembly-line workers are more fatalistic, less sympathetic with and tolerant of others. The line workers are also more radical in their political views.

Recent investigations of Arthur Kornhauser tend to confirm these findings. In an as yet unpublished work and in a recent article, Kornhauser explores the relationships between skill level and various psychological and attitude variables. He finds that skill level in manual jobs is clearly related to mental health and also to certain political and social attitudes. He concludes that the higher the skill level of the job, the more likely that men will show high mental health, while mental health is lowest among workers with repetitive conveyor-paced jobs.[6]

Kornhauser's measures of mental health coincide, in part, with dimensions of attitudes also examined in this study. His indices include measures of: anxiety and tension; hostility versus trust in people; self-esteem versus negative self-feelings; sociability versus withdrawal; personal morale versus anomie or social alienation. He also found, consistent with the findings here, that workers in lower-skill jobs were higher on economic 'liberalism', but lower in liberal attitudes on race relations and international affairs. Kornhauser concludes that the workers at various skill levels differ most sharply in feelings of defeat, pessimism, personal inadequacy, futility and distrust of others.[7]

Assembly-line workers, as we have seen, do not find their work interesting. They complain bitterly and consistently about monotony. In many respects, automobile assembly-line work reflects the most unhealthy characteristics of industrial work. Line workers have a sense of futility. They are frustrated and at the same time are seeking to undo their sense of inadequacy. Their frustrations are clearly reflected in their alienation, i.e. their feelings of powerlessness, and its attendant consequences. Their feelings of helplessness are reflected in their fatalistic acceptance of their lot and of the agonies of the great world. Their efforts to restore their self-esteem are reflected in their desire for economic improvement.

What are the implications of these findings for our understanding of political life?

First, it seems reasonable to suppose that assembly-line workers and others whose jobs deprive them of control over their physical movements are more likely to sympathise with radical and mass agitations. Assembly-line workers are more isolated and more likely to feel disenfranchised than other industrial workers. Their feelings about politics are likely to be influenced by desires for relatively radical change, even if these feelings are partly suppressed.

Second, it seems clear that the job structure of industry is one factor that needs to be understood in grasping the impact of industrialisation. Kornhauser has pointed out that it is not industrialisation *per se*, but rather 'discontinuities' in industrial development, that have been the basis for anti-democratic movements. Among these discontinuities are excessive rapidity of industrialisation or the acute breakdown of the economy. Perhaps along with these factors we should also turn our attention to the job structure of emerging industries, focusing on the proportion of assembly-line type jobs.

Third, specific job groups may show a clearer relationship to political party allegiance in countries with multi-party systems than they seem to in America. The great majority of workers in all of the sample groups examined here were Democrats, yet the groups differed in the extent to which they favoured liberal economic measures, and in the intensity of their antagonism to the Republicans. In a country where a greater variety of political choices is offered, such differences within the strata of manual workers may show up more sharply in party allegiances. And conversely, if most of the men in most circumstances vote Democratic anyway, regardless of differences in their feelings, party organisers here may well be content to get them to the polls and ignore their feelings.

Finally, how widely can these findings be generalised? If recent work in industrial psychology is any indication, a large percentage of manual and white-collar workers are not interested in their work. It seems as yet an open question whether interest in work and control over work will increase or decrease with the spread of various forms of automation. At any rate, millions, if not tens of millions of Americans in the foreseeable future are likely to be working at uninteresting, repetitive jobs. Perhaps such workers, even if they are not manual workers, will internalise the tensions and dissatisfactions of their work.

I am not attempting to argue that a bored clerk in a department store, a dish-washer in a restaurant, an auto assembly worker and an I.B.M. card puncher are likely to respond to their jobs in the same manner. Their political and social attitudes are conditioned by the content and status of their jobs, as well as the circumstances and expectations they bring with them, and other factors. Yet it seems possible that at each status level, those with the most repetitive, least interesting or controllable jobs will be the most dissatisfied, alienated and politically 'radical'.

VI

Various political analysts have argued that there is a causal connection between industrialisation and democracy. Among the most prominent of these probably is Lipset, who has shown correlations between stable democratic systems and various indices which provide a way of measuring the relative modernity of nations. In his view, contradicting Marx, industrialisation tends to produce the preconditions for political democracy because the class struggle is shaped in moderate directions. Relative wealth and increased education tend to preclude extremism in the lower classes. The absence of deep and stable class differences also makes it possible for the rich and privileged to acknowledge the poor as members of the same species.

Robert Lane has recently buttressed Lipset's argument from another direction. His research, in depth interviews, of the attitudes of a sample of 'common men' in an Eastern city reveals that the work situation in which these men find themselves is conducive to the formation of democratic norms. Most of his sample have substantial independence at work, find satisfaction in their jobs, exercise some skill, are in a position to assert their competence and to exchange views with co-workers and superiors. Lane's findings also move in the direction of substantiating Blauner's hypothesis that modern industrial work provides wider opportunities for the exercise of skill than work in previous societies. Lane concludes that the data on industrial work substantiate Lipset's correlation between democracy and advanced industrial development.

The findings reported here, as well as those of Kornhauser noted above, are not nearly so optimistic as Lane's or Blauner's.

It seems that at some lower skill levels work alienation is considerable, and has serious psychological and political consequences, some of which are not very healthy for a democratic polity. On the basis of the very incomplete knowledge we have about the effects of different types of work, scepticism about the overall consequences of industrial work is still in order.

The manual workers examined in this study are not extremists in the sense of favouring truly radical change, revolution, violence or of lending support to the undemocratic political movements. But it is not their jobs, at least in the case of the two groups of semi-skilled workers, that have made them supporters of the *status quo*, nor taught them very much about the nature of democratic social relationships. If these men are defenders of the political *status quo* it seems to be largely because of their status as consumers, not because of their work lives. If they have very much sense of personal adequacy, it is because of the control and judgement they are able to exercise outside their jobs. Though industrial work, for many reasons, may be more conducive to the development of democratic sentiments than the life of the peasant, it is quite another thing to argue that industrial work, at least of the repetitive variety, is generally healthy or enhances an individual's sense of worth.

Lipset's argument concerning the relationship between industrialisation and democracy seems to run counter to the analyses of writers such as C. Wright Mills and Erich Fromm. Fromm and Mills emphasise the alienation and apathy characteristic of bureaucratised industrial societies. They argue that democracy is far from being a reality in today's world. As they picture it, many ordinary citizens have become passive and alienated, feel unable to control their fates, and have lost sight of the significant issues of political life. Both Mills and Fromm relate these characteristics of industrial societies to the structure of work. Both argue that alienation and passivity at work breed alienation and passivity elsewhere. If they are right and industrialisation creates many kinds of work situations which are not conducive to the development of democratic attitudes, then how can industrialisation and democracy be related?

A great part of the difficulty in resolving this antithesis may lie in definitions and points of view. Lipset's definition of democracy is an arbitrary, operational one. He speaks of two characteristics: competitive parties and the absence of strong anti-democratic political groups. It is a 'democracy' fitting this definition which Lipset finds related to his indices of modernity. Both Fromm and Mills speak of democracy and democratic attitudes in a different sense. Both bring normative judgements to bear on present reality. So, it is possible to agree that Lipset's correlation is correct and important, and yet to agree with Mills and Fromm that present-day democracies are not as democratic as they might be.

Robert Lane has argued that ego strength is a crucial psychological correlate of democracy. Without having experienced an inner sense of mastery as well as a sense of mastery over the environment, men do not develop the requisite ego strength to pursue a consistent, long-term course, or to effect much social change. One of the most important, if unproven, implications of this study is that the work men do can have important consequences for their ego strength. Here the concerns of Fromm and Mills are significant. Men in a work environment they cannot control may be to some unknown degree damaged in their sense of mastery, and this damage may render them less capable of coping with and altering an environment they find unsatisfactory. To the extent that such damage occurs, men become victims rather than creators. To the extent that work life contributes to such incapacities, it needs improvement in modern societies. Industrial democracy may yet prove to be a prerequisite to political and social democracy.

A few final *caveats*: first, the severe dissatisfactions found among automobile assembly-line workers are probably not typical of the attitudes of most modern industrial workers. The nature of the assembly-line itself intensifies the problems of a mechanised job. These findings therefore represent only one segment of the working class. Their relevance to men in other types of jobs which are less intensely disliked remains to be seen. Second, the findings presented in the tables above are for the most part based on a rather crude dichotomising of attitudes as these attitudes appeared in the setting of open-ended interviews. Such findings, especially in view of the small size of the sample, are at best suggestive and exploratory. Most important, the question of the intensity with which various attitudes are held must remain moot in many of the areas this article has explored. For example, though many of the men criticise what they consider to be manipulative

practices by political figures, most of them are not 'alienated' from the political system. How deep their cynicism may go on occasion, is a question this study has not touched.

Notes

1. The extent of satisfaction expressed is clearly in part a function of the particular question asked. For example, a group of American unskilled workers asked 'Are you satisfied or dissatisfied with your present job?' showed 72 per cent indicating they were satisfied. Exactly what satisfaction means in such a case is not at all clear.

2. Generally speaking, these findings confirm previous ideas about working-class political attitudes in the United States.

3. The Liberalism–Conservatism Scale is taken from Richard Centers's book *The Psychology of Social Classes* (Princeton: Princeton University Press, 1949). The scale was part of the questionnaire administered with each interview. One man in the repair category never completed this questionnaire and so only 13, instead of 14 men are scored above. The Centers scale basically compares nineteenth-century liberalism (conservative) with twentieth-century liberalism (radical – welfare state). The questions explore management versus labour identification; belief in the possibilities of economic opportunity; attitude towards government ownership and social insurance.

4. Lipset points out that the unemployed, feeling resentful and alienated, are less tolerant of minority groups than are employed workers, a finding that seems related to the conclusions here about the assembly-line group.

5. Thus, most of these men are not 'cabalists' as that term is used by Robert E. Lane. Describing the cabalist frame of mind Lane notes: (1) there is some unofficial, quasi-conspiratorial group behind the scenes to manipulate and control public affairs; (2) each cabal group is responsible to no one but itself; and (3) the cabalist argument is protean; for the same person it will focus now upon the international banker and now upon the Communists. . . .

Lane argues that such a pattern of thought is rooted both in personality weaknesses and the social processes of modern society. He finds this frame of mind among men with anti-democratic tendencies.

6. The following table will indicate his findings:

MENTAL HEALTH OF FACTORY WORKERS

Men in 40s (age)	*% with good mental health scores*	*N*
Skilled	56	45
High semi-skilled	41	98
Ordinary semi-skilled	38	82
Repetitive semi-skilled	26	73
Repetitive machine-paced only (subdivision of preceding category)	16	32

Source: A. Kornhauser, 'Mental Health of Factory Workers: A Detroit Study', *Human Organisation*, vol. 21 (1962) 45.

These relationships held when the author controlled for education.

7. Though I have spoken of a causal relationship between jobs and attitudes, I have not really 'proven' such a relationship exists. The crucial problem is one of pre-selection. Perhaps men on assembly-line jobs differ from other semi-skilled workers because they were different before they ever took these jobs, rather than because they have been changed by their employment. Perhaps it is only men with a low sense of self-esteem and strong withdrawal tendencies who stay on the assembly-line. Kornhauser tries to handle this problem by exploring the life of his interviewees before they took their present jobs. He concludes that there is demonstrable evidence that jobs change attitudes, regardless of what those attitudes were before the job was taken. If so, then personality differences that exist before taking the job do not explain occupational group differences in mental health.

The study of government in society has tended to concentrate on the State, yet government and politics exist and operate in other associations than the State. Not only political institutions like parties and interest groups but also football clubs, like that at Pentrediwaith, or other associations are microcosms of the State with their own government and internal politics. This internal politics can have a significance for the State and can be studied for this reason, but it can also operate in a way that has little if any significance for the State while nevertheless having an importance for those who are seeking simply to understand government and politics as integral features of the life of man in society.

The first of the two extracts in this chapter comes from Robert Michels's book 'Political Parties', first published in German in 1911. This book broke new ground in the social sciences. It was a study of the internal working of democratic parties in Europe, and Michels argued that there were common trends towards oligarchy in party life. His study centred on the leadership in the parties, and the extract below describes one feature of the leaders' behaviour and at the same time shows the working of politics within the parties. This passage by Michels is complemented by the extract from R. Frankenberg's 'Village on the Border', a study of life in a Welsh border village, which gives an account of politics within an institution that is not immediately involved in the politics of the State.

THE STRUGGLE AMONG THE LEADERS THEMSELVES

Robert Michels

THE thesis of the unlimited power of the leaders in democratic parties requires, however, a certain limitation. Theoretically the leader is bound by the will of the mass, which has only to give a sign and the leader is forced to withdraw. He can be discharged and replaced at any moment. But in practice, as we have learned, for various reasons the leaders enjoy a high degree of independence. It is none the less true that if the democratic party cannot dispense with autocratic leaders, it is at least able to change these. Consequently the most dangerous defect in a leader is that he should possess too blind a confidence in the masses. The aristocratic leader is more secure than the democratic against surprises at the hands of the rank and file. It is an essential characteristic of democracy that every private carries a marshal's baton in his knapsack. It is true that the mass is always incapable of governing; but it is no less true that each individual in the mass, in so far as he possesses, for good or for ill, the qualities which are requisite to enable him to rise above the crowd, can attain to the grade of leader and become a ruler. Now this ascent of new leaders always involves the danger, for those who are already in possession of power, that they will be forced to surrender their places to the new-comers. The old leader must therefore keep himself in permanent touch with the opinions and feelings of the masses to which he owes his position. Formally, at least, he must act in unison with the crowd, must admit himself to be the instrument of the crowd, must be guided, in appearance at least, by its goodwill and pleasure. Thus it often seems as if the mass really controlled the leaders. But whenever the power of the leaders is seriously threatened, it is in most cases because a new leader or a new group of leaders is on the point of becoming dominant, and is inculcating views opposed to those of the old rulers of the party. It then seems as if the old leaders, unless they are willing to yield to the opinion of the rank and file and to withdraw, must consent to share their power with the new arrivals. If, however, we look more closely into the matter, it is not difficult to see that their submission is in most cases no more than an act of foresight intended to obviate the influence of their younger rivals. The submission of the old leaders is ostensibly an act of homage to the crowd, but in intention it is a means of prophylaxis against the peril by which they are threatened – the formation of a new elite.

The semblance of obedience to the mass which is exhibited by the leaders assumes, in the case of the feebler and the more cunning among them, the form of demagogy. Demagogues are the courtesans of the popular will. Instead of raising the masses to their own level, they debase themselves to the level of the masses. Even for

the most honest among them, the secret of success consists in 'knowing how to turn the blind impulsiveness of the crowd to the service of their own ripely pondered plans'.[1] The stronger leaders brave the tempest, well knowing that their power may be attacked, but cannot be broken. The weak or the base, on the other hand, give ground when the masses make a vigorous onslaught; their dominion is temporarily impaired or interrupted. But their submission is feigned; they are well aware that if they simply remain glued to their posts, their quality as executants of the will of the masses will before long lead to a restoration of their former dominance. One of the most noted leaders of German socialism said in a critical period of tension between the leaders and the masses, that he must follow the will of the masses in order to guide them.[2] A profound psychological truth is hidden in this sarcasm. He who wishes to command must know how to obey.

It has been affirmed that popular revolutions usually end by destroying their leaders. In proof there have been quoted the names of Rienzi, Masaniello and Michele di Lando, for Italy, and of Danton and Robespierre, for France. For these and many similar instances the observation is a true one. It would, however, be an error to accuse the crowd of rising against its leaders, and to make the masses responsible for their fall. It is not the masses which have devoured the leaders: the chiefs have devoured one another with the aid of the masses. Typical examples are that of Danton, who was overthrown by Robespierre, and that of Robespierre, who was destroyed by the surviving Dantonists.

The struggle which arises between the leaders, and their mutual jealousies, induce them to employ active measures and often to have recourse to artifices.[3] Democratic deputies endeavour to disarm their adversaries within the party, and at the same time to acquire a new prestige in the eyes of the masses, by displaying in parliament 'a formidable activity on behalf of the common cause'. This is regarded at once as a democratic duty and as a measure of personal precaution. Since the great majority of the deputies, electors and comrades have no precise ideas concerning the functions he exercises, and are continually inclined to accuse him of slackness, the deputy is from time to time forced to recall himself to their memories.[4] It is this need which has given rise to not a few of those speeches to which the Germans give the name of *Dauerreden* (interminable speeches), and it has also been the cause of more than one 'scene' in the various parliaments of Austria, France, England and Italy. It is, in fact, held that the most efficacious means for retaining the attention of the masses and of rendering them proud of their leaders is to be found in the provocation of those personal incidents which are far more interesting to the great public and far more within the scope of its intelligence than a report upon the utilisation of water power or upon a commercial treaty with the republic of Argentina. Moreover, it has to be remembered that in many countries, and above all in Italy, such scenes are recorded in the capitalist press with the greatest abundance of detail, while serious speeches are summed up in a few lines, and with especial brevity when the speaker is a socialist. Thus even in normal times the oratorical activity of the parliamentary representatives of the democratic parties is considerable. In Italy, the socialist deputies have boasted that between 25 March and 10 July 1909, they spoke in the Chamber 212 times. The figure represents 20·4 per cent of all the speeches made in parliament during the period, while the socialist deputies at this time constituted only 8 per cent of the members.[5] Such loquacity serves not merely to maintain the prestige of the party in the eyes of its opponents, but is also a matter of personal interest to each deputy, being a means to secure his re-election in competition, not only with enemies in other parties, but also with jealous rivals belonging to his own organisation.

The differences which lead to struggles between the leaders arise in various ways. Reference has previously been made to the inevitable antagonism between the 'great men' who have acquired a reputation in other fields, and who now make adhesion to the party, offering it their services as generals, and the old-established leaders, who have been socialists from the first.[6] Often conflict arises simply between age and youth. Sometimes the struggle depends upon diversity of social origin, as when there is a contest between proletarian leaders and those of bourgeois birth.[7] Sometimes the difference arises from the objective needs of the various branches of activity into which a single movement is subdivided, as when there is a struggle between the political socialist party and the trade union element, or within the political party between the parliamentary group and the executive. In some cases there is a horizontal stratification, causing a struggle

between one stratum of the bureaucracy and another; at other times the stratification is vertical, as when there occurs a conflict between two local or national groups of leaders; between the Bavarian socialists and the Prussian; between those of Frankfurt and those of Hanau; between the French followers of Vaillant, Jaurès and Hervé, and the German adherents of Bebel and von Vollmar (in the anti-militarist discussion at the international congress of Stuttgart). Often enough struggles among the socialists are the outcome of racial differences. The unceasing contests in the international congresses between the German socialists and the French afford in more than one respect a parallel with the Franco-German War of 1870. In these same congresses there participates a third group, misunderstood and heterogeneous, the representatives of English socialism, hostile to all the others and encountering the enmity of all. In most cases, however, the differences between the various groups of leaders depend upon two other categories of motives. Above all there are objective differences and differences of principle in general philosophical views, or at least in the mode in which the proximate social evolution is conceived, and consequent divergences of opinion as to the most desirable tactics: this leads to the manifestation of the various tendencies known as reformist and Marxist, syndicalist and political socialist and so on. In the second place, we have the struggles that depend on personal reasons: antipathy, envy, jealousy, a reckless attempt to grasp the first positions, demagogy. Enrico Ferri said of his opponent Filippo Turati: 'He hates me because he thinks there is not room for two cocks in the same fowl-house.'[8] In most cases the two series of motives are somewhat confounded in practice; and in the long run we find that those of the former series tend to be replaced by those of the latter, inasmuch as differences of principle and of the intellectual order soon become personal and lead to a profound hostility between the representatives of the various theories. Conversely it is clear that motives of the second series, since those who are influenced by them are ashamed to display them in their true colours, always endeavour to assume the mantle of theory; personal dislike and personal hostility pompously masquerade as differences of views and tactics.

The oligarchy which issues from democracy is menaced by two grave dangers: the revolt of the masses, and (in intimate relationship with this revolt, of which it is often the result) the transition to a dictatorship when one among the oligarchs succeeds in obtaining supreme power. Of these two dangers, one comes from below, while the other arises within the very bosom of the oligarchy: we have rebellion on one side, and usurpation on the other. The consequence is that in all modern popular parties a spirit of genuine fraternity is conspicuously lacking; we do not see sincere and cordial mutual trust; there is a continual latent struggle, a spirit of irritation determined by the reciprocal mistrust of the leaders, and this spirit has become one of the most essential characteristics of every democracy. The mistrust of the leaders is directed above all against those who aspire to command their own organisations. Every oligarchy is full of suspicion towards those who aspire to enter its ranks, regarding them not simply as eventual heirs but as successors who are ready to supplant them without waiting for a natural death. Those who have long been in possession (and this applies just as much to spiritual and psychical possession as to material) are proud of their past, and are therefore inclined to look down upon those whose ownership is of more recent date. In certain Sicilian towns, struggles go on between two parties who in popular phrase are ironically termed *i ricchi* and *gli arricchiti* (the wealthy and those who have attained to wealth). The former consists of the old landed gentry; while the latter, the parvenus, are merchants, contractors for public works, manufacturers and the like.[9] A similar struggle makes its appearance in modern democratic parties, although it is not in this case characterised by any flavour of economic distinction. Here also we have a struggle between the *détenteurs d'emploi et les chercheurs d'emploi,* or as the Americans put it, between the 'ins' and the 'outs'. The latter declare war on the former, ostensibly on the ground of eternal principle, but in reality, in most cases, because in such opposition they find the most effective means of forcing their way into the circle of the chiefs. Consequently in meetings they display themselves as implacable theoretical adversaries, 'talking big' solely in order to intimidate the accepted leaders, and in order to induce them to surrender a share of the spoil to these turbulent comrades. Often enough, the old leaders resist, and maintain their ground firmly; in such cases their opponents, changing front, abandon the attitude of struggle, and attach themselves to the triumphal car of the men in power,

hoping thus to attract favour, and, by a different route, to realise their own ambitions.[10]

The struggle between the old leaders and the aspirants to power constitutes a perpetual menace to freedom of speech and thought. We encounter this menance in every democratic organisation in so far as it is well ordered and solidly grounded, and in so far as it is operating in the field of party politics (for in the wider life of the State, in which the various parties are in continual reciprocal concussion, it is necessary to leave intact a certain liberty of movement).[11] The leaders, those who already hold the power of the party in their hands, make no concealment of their natural inclination to control as strictly as possible the freedom of speech of those of their colleagues from whom they differ.

The consequence is that those in office are great zealots for discipline and subordination, declaring that these qualities are indispensable to the very existence of the party. They go so far as to exercise a censorship over any of their colleagues whom they suspect of rebellious inclinations, forcing them to abandon independent journals, and to publish all their articles in the official organs controlled by the leaders of the majority in the party. The prohibition, in the German socialist party, of collaboration on the part of its members with the capitalist press, is in part due to the same tendency; while the demand that the comrades should have nothing to do with periodicals which, though socialist, are founded with private capital and are not subject to the official control of the party executive, arises solely from this suspicion on the part of the leaders.[12]

In the struggle against the young aspirants, the old leader can as a rule count securely upon the support of the masses. The rank and file of the working-class parties have a certain natural distrust of all new-comers who have not been openly protected or introduced into the party by old comrades; and this is above all the case when the new-comer is derived from another social class. Thus the new recruit, before he can come into the open with his new ideas, must submit, if he is not to be exposed to the most violent attacks, to a long period of quarantine. In the German socialist party, this period of quarantine is especially protracted, for the reason that the German party has been longer established than any of the others, and because its leaders therefore enjoy an exceptional prestige. Many of them were among the actual founders of the party, and their personalities have been consecrated by the baptism of fire which they suffered during the enforcement of the anti-socialist laws. A socialist who has had his party card in his pocket for eight or ten years is often regarded in his branch as a 'young' member. This tendency is reinforced by the respect for age which is so strong among the Germans, and by the tendency towards hierarchy of which even the democracy has not been able to divest itself. Finally, it may be added that the bureaucracy of the German labour movement, like every strongly developed bureaucracy, tends instinctively towards exclusivism. Consequently in the German social democracy, in contradistinction to other socialist parties which are less solidly organised, we find that not merely the recently enrolled member of the party (the so-called *Fuchs*), but also the ordinary member who does not live in the service and by the service of the party but has preserved his outward independence as a private author or in some other capacity, and has therefore not been incorporated among the cogwheels of the party machine, very rarely succeeds in making his influence felt. There can be no doubt that this fact plays a large part in the causation of that lack of a number of capable young men, displaying fresh energies, and not greatly inferior to the old leaders, a lack which has often been deplored. The annual congresses of the socialist party have even been spoken of as 'congresses of the party officials'. The criticism is not unjust, for among the delegates to the socialist congresses the percentage of party and trade union officials is enormous.[13] It is above all in the superior grades of the organisation that the tendencies we are here analysing are especially conspicuous. In Germany, the management of the socialist party is not entrusted to young men, as often happens in Italy, or to free publicists, as in France, but to old members, *des anciens*, elderly officials of the party. Moreover, the conservative psychology of the masses supports the aspirations of the old leaders, for it would never occur to the rank and file to entrust the care of their interests to persons belonging to their own proper sphere, that is to say, to those who have no official position in the party and who have not pursued a regular bureaucratic career.[14]

Often the struggle between the old leaders in possession of power and the new aspirants assumes the aspects of a struggle between responsible and irresponsible persons.[15] Many criticisms

levelled by the latter against the former are beside the mark, because the leaders have grave responsibilities from which the aspirants are free. This freedom gives the aspirants a tactical advantage in their conflict with the old leaders. Moreover, precisely because they are irresponsible, because they do not occupy any official position in the party, the opponents are not subject to that simulacrum of democratic control which must influence the conduct of those in office.

In order to combat the new chiefs, who are still in a minority, the old leaders of the majority instinctively avail themselves of a series of underhand methods through which they often secure victory, or at least notably retard defeat. Among these means, there is one which will have to be more fully discussed in another connection. The leaders of what we may term the 'government' arouse in the minds of the masses distrust of the leaders of the 'opposition' by labelling them incompetent and profane, terming them spouters, corrupters of the party, demagogues and humbugs, while in the name of the mass and of democracy they describe themselves as exponents of the collective will, and demand the submission of the insubordinate and even of the merely discontented comrades.

In the struggle among the leaders an appeal is often made to loftier motives. When the members of the executive claim the right to intervene in the democratic functions of the individual sections of the organisation, they base this claim upon their more comprehensive grasp of all the circumstances of the case, their profounder insight, their superior socialist culture and keener socialist sentiment. They often claim the right of refusing to accept the new elements which the inexpert and ignorant masses desire to associate with them in the leadership, basing their refusal on the ground that it is necessary to sustain the moral and theoretical level of the party. The revolutionary socialists of Germany demand the maintenance of the centralised power of the executive committee as a means of defence against the dangers, which would otherwise become inevitable as the party grows, of the predominant influence of new and theoretically untrustworthy elements. The old leaders, it is said, must control the masses, lest these should force undesirable colleagues upon them. Hence they claim that the constituencies must not nominate parliamentary candidates without the previous approval of the party executive.[16]

The old leaders always endeavour to harness to their own chariot the forces of those new movements which have not yet found powerful leaders, so as to obviate from the first all competition and all possibility of the formation of new and vigorous intellectual currents. In Germany, the leaders of the socialist party and the trade union leaders at first looked askance at the Young Socialist movement. When, however, they perceived that this movement could not be suppressed, they hastened to place themselves at its head. There was founded for the guidance of the socialist youth a 'Central Committee of Young German Workers', comprising four representatives from each of the three parties, that is to say, four from the executive of the socialist party, four from the general committee of trade unions and four from the Young Socialists (the representatives of the latter being thus outnumbered by two to one).[17] The old leaders endeavour to justify the tutelage thus imposed on the Young Socialists by alleging (with more opportunist zeal than logical acuteness) the incapacity of the youthful masses, if left to their own guidance, of wisely choosing their own leaders and of exercising over these an efficient control.[18]

We have by no means come to an end of our enumeration of the weapons at the disposal of the old leaders in their conflict with the new aspirants to power. Charlemagne effected the final subjugation of the Saxon tribal chiefs by making them counts. In this way he not only increased the brilliancy of their position, but also gave them a restricted share in his own power. This means has been practised again and again in history, where an old ruler has wished to render harmless, insubordinate but influential chiefs, and thus to prevent a rebellion against his own authority. Oligarchies employ this stratagem with just as much success as monarchies. The feudal state of Prussia appointed to the privy council the most defiant among the leaders of its bourgeoisie. At a time when the youthful German bourgeoisie was still filled with a rebellious spirit towards the nobility and towar ls the traditional authority of the State, this tendency aroused much bitterness. Thus Ludwig Börne wrote in 1830:

> Wherever a talented force of opposition has made itself apparent and has secured respect from those in authority, it is chained to the professorial chair, or is controlled by being

> harnessed to the government. If the governmental ranks are full, so that no place can be found for the new energies, a state livery is at least provided for the authors by giving them titles and orders. In other cases the dangerous elements are isolated from the people by immuring them in some noble's castle or princely court. It is for this reason that nowhere else do we find so many privy councillors as in Germany, where the courts are least inclined to take any one's advice.[19]

In the Spanish elections of 1875, we learn that so great was the popular indifference that the government had matters altogether in its own hands, but in order to be secure in any event it thoughtfully selected a certain number of opposition candidates.[20] It seems that things are much the same in Spain even today.[21] These tactics are not confined to States that are still permeated by feudal conceptions. Where plutocratic rule is supreme, corruption persists unchanged and it is only the corrupter who is different. This is plainly shown by Austin Lewis when he writes:

> The public ownership contingent in politics being composed of the middle and subjugated class have neither the political ability nor the vital energy necessary for the accomplishment of the task which they have undertaken. The brains of the smaller middle class have already been bought by the greater capitalists. Talent employed in the service of the chiefs of industry and finance can command better prices than can be obtained in the uncertain struggle for economic standing which members of the middle class have to wage. The road to professional and political preferment lies through the preserves of the ruling oligarchy, whose wardens allow no one to pass, save servants in livery. Every material ambition of youth is to be gratified in the service of the oligarchy, which shows, generally, an astuteness in the selection of talent that would do credit to a bureaucrat or a Jesuit.[22]

Of late years the ruling classes in the countries under a democratic regime have hoped to impose obstacles in the way of the revolutionary labour movement by conceding posts in the ministry to its most conspicuous leaders, thus gaining control over the revolutionary impulse of the proletariat by allowing its leaders to participate in power, though cautiously and in an extremely restricted measure. The oligarchy which controls the modern democratic party has often employed the same means to tame the opposition. If the leaders of the opposition within the party are dangerous because they have a large following among the masses, and if they are at the same time few in number, the old party leaders endeavour to hold them in check and to neutralise their influence by the conciliatory methods just described. The leaders of the opposition receive high offices and honours in the party, and are thus rendered innocuous – all the more so seeing that they are not admitted to the supreme offices, but are relegated to posts of the second rank which give them no notable influence, and they are without hope of one day becoming a majority. On the other hand, they divide with their ancient adversaries the serious weight of responsibility which is generated by common deliberations and manifestations, so that their activities become confounded with those of the old leaders.[23]

In order to avoid having to divide their power with new elements, especially such as are uncongenial by tendency or mental characteristics, the old leaders tend everywhere with greater or less success to acquire the right of choosing their own colleagues, thus depriving the masses of the privilege of appointing the leaders they themselves prefer.[24]

The path of the new aspirants to power is always beset with difficulties, bestrewn with obstacles of all kinds, which can be overcome only by the favour of the mass. Very rarely does the struggle between the old leaders and the new end in the complete defeat of the former. The result of the process is not so much a *circulation des élites* as a *réunion des élites*, an amalgam, that is to say, of the two elements. Those representing the new tendency, as long as their footing is still insecure, seek all sorts of side paths in order to avoid being overthrown by the powers-that-be. They protest that their divergence from the views of the majority is trifling, contending that they are merely the logical advocates of the ancient and tried principles of the party, and express their regret that the old leaders display a lack of true democratic feeling. Not infrequently it happens that they avert the blows directed against them by craftily creeping behind the backs of their established and powerful opponents who are about to annihilate them, solemnly declaring, when wrathful blows are directed against them, that they are in complete accord with the

old leaders and approve of all their actions, so that the leaders seem to be beating the air. On many occasions in the recent history of the socialist parties, the reformist minorities, in order to avoid destruction, have bowed themselves beneath the yoke of the so-called revolutionary majorities by voting (with a fine practical and tactical sense, but with an entire lack of personal pride and political loyalty) resolutions which were drafted precisely in order to condemn the political views dear to the minority.'[25] In two cases only does it sometimes happen that the relationships between the two tendencies become strained to the breaking-point. In the first place this may happen when the leaders of one of the two factions possess a profound faith in their own ideas, and are characterised at once by tactical fanaticism and theoretical irreconcilability – or, in other words, when the objective reasons which divide them from their opponents are felt with an unaccustomed force and are professed with an unwonted sincerity. In the second place it may happen when one of the parties, in consequence of offended dignity or reasonable susceptibility, finds it psychologically impossible to continue to live with the other, and to carry on within the confines of the same association a continued struggle for dominion over the masses. The party will then break up into two distinct organisms, and in each of these there will be renewed the oligarchical phenomena we have been describing.

One of the most interesting chapters in the history of the struggles between leaders deals with the measures which these leaders adopt within their own closed corporations in order to maintain discipline – that is to say, in order to preserve the cementing force of the will of the majority. In the struggle which the various groups of leaders carry on for the hegemony of the party, the concept of democracy becomes a lure which all alike employ. All means are good for the conquest and preservation of power. It is easy to see this when we read the discussions concerning the system to be employed for the appointment of the party executive. The various tendencies manifested in this connection all aim at the same end, namely, at safeguarding the dominance of some particular group. Thus in France the Guesdists, whose adherents are numerous but who control a small number only of the groups, advocate a system of proportional representation; the Jauressists, on the other hand, who are more influential in respect of groups than of members, and also the Hervéists, oppose proportional representation within the party, for they fear that this would give the Guesdists group too great a facility for the enforcement of its own special methods of action, and they propose to maintain the system of local representation or of representation by delegation.[26]

In the American Congress, each party possesses a special committee which exercises a control over the attendance of its members at the sessions, and which on the occasion of decisive votes issues special summonses or 'whips'. When an interesting bill is before the house, the party committee also summons a *caucus*, that is to say, a private meeting of the parliamentary group, and this decides how the congressmen are to vote. All members of the party are bound by the decision of such a caucus. Naturally no immediate punishment is possible of those who rebel against the authority of the caucus; but at the next election the independent congressman is sure to lose his seat, for the party managers at Washington will not fail to report to their colleagues, the bosses of the local constituency, the act of insubordination committed by the congressman concerned. The most vital of all the caucuses is that which precedes the election of the speaker of the congress. The ideas and sympathies of the speaker have a decisive influence upon the composition of the committees and therefore upon the whole course of legislation. For this reason his election is of fundamental importance, and is preceded for several weeks by intrigues and vote-hunting campaigns. Doubtless it is not in every case that the votes are decided in advance at a meeting of the group. Where laws of minor importance are concerned, every member of Congress is free to vote as he pleases. But in times of excitement obedience is exacted, not only to the decisions of the caucus, but also to the authority of the party leaders. This last applies especially to Congress, for in the Senate the members are extremely jealous of their absolute equality. On the other hand, the caucus has an even greater importance in the case of the Senate, for here the groups are smaller and the caucus can therefore function more efficiently. The groups in Congress may number more than two hundred members, whereas those of the Senate rarely exceed fifty.[27]

The parliamentary group of the German social democracy is likewise dominated, as far as its internal structure is concerned, by a most rigorous application of the principle of subordination.

The majority of the parliamentary group decides the action of all its members on the various questions submitted to the Reichstag or to the diets, exercising what is known as the *Fraktionszwang* (group coercion). No individual member has the right to independent action. Thus the parliamentary group votes as a single entity, and this not merely in questions of a distinctively socialist bearing, but also in those which are independent of socialist ideas, and which each might decide according to his own personal conceptions. It was very different in the French parliament during the fratricidal struggle between the Jauressists and the Guesdists before the attainment of socialist unity in France, for at that time each deputy used to vote as he pleased. But the German example shows that liberty of opinion no longer exists where the organisation demands common action and where it has some force of penetration in political life.

In certain cases, however, all these preventive measures fail of their effect. This happens when the conflict is not simply between a minority and a majority within the group, but between the group and one single member who possesses outside parliament, in certain sections of the party, the full support of the subordinate leaders. When a conflict occurs in such conditions, the deputy, though isolated, is sure of victory. The electors, in fact, usually follow with great docility the oscillations and evolutions of their parliamentary representatives, and they do this even in constituencies where socialist voters predominate. The ministers Briand, Viviani and Millerand have been expelled from the French socialist party, but the former members of the socialist organisations in their constituencies have remained faithful to these leaders, resigning from the socialist party, and continuing as electors to give the ex-socialists their support. Analogous were the cases of John Burns in England (Battersea) and of Enrico Ferri in Italy (Mantua). It was enough in Ferri's case that at an appropriate moment he should reveal a new truth to produce immediately a collective change in the political opinions of an entire region. Having first been, with Ferri, revolutionary and irreconcilable, this region became converted in a single night, always following Ferri, to the principle of class co-operation and of participation in ministerial activity.[28] In Germany, the party executive had to make use of all its authority in order, at the last minute, to induce the comrades of Chemnitz to withdraw their support from their deputy Max Schippel, and those of Mittweida from Otto Göhre, when these two deputies had displayed heterodox leanings.

The tendency of the deputy to set himself above his party is most plainly manifest precisely where the party is strongly organised; especially, therefore, in the modern labour parties; and within these, again, more particularly in the reformist sections. The reformist deputies, as long as they have not upon their side a majority within the party, carry on an unceasing struggle to withdraw themselves from the influence of the party, that is to say, from the mass of the workers who are organised as a party. In this period of their evolution they transfer their dependence upon the organised mass of the local socialist section to the electors of the constituency, who constitute a grey, unorganised and more or less indifferent mass. Thus from the organised masses, who may be under the influence of their opponents within the party, they appeal to the mass of the electors, with the contention that it is to these latter alone, or at least chiefly, that they have to give an account of their political conduct. It is right to recognise that this appeal to the electorate as the body which has conferred a political mandate is frequently based upon genuinely democratic sentiments and principles. Thus, at the international socialist congress of London (1893), the four French socialist deputies refused to make use of the mandates which had been conferred upon them by political or corporative groups, thus defying the rules of admission to the congress. After extremely violent discussions they were ultimately admitted simply as deputies, having raised the question of principle whether an important constituency capable of returning a socialist deputy to the Chamber should not have the same rights which are granted to a local socialist or trade union branch, especially when it is remembered that such a branch may consist of a mere handful of members.[29] It is true that in certain circumstances a constituency inspired by socialist sentiment, even if it be not socialistically organised, constitutes a better basis, in the democratic sense, for political action than a small socialist branch whose members are mostly petty bourgeois or lawyers;[30] and even if a large local organisation exists, the constituency as a whole is a better basis than a badly attended party meeting for the selection of a candidate.[31]

From our study of the intricate struggles which proceed between the leaders of the majority and those of the minority,

between the executive organs and the masses, we may draw the following essential conclusions.

Notwithstanding the youth of the international labour movement, the figures of the leaders of that movement are more imposing and more imperious than those displayed in the history of any other social class of modern times. Doubtless the labour movement furnishes certain examples of leaders who have been deposed, who have been abandoned by their adherents. Such cases are, however, rare, and only in exceptional instances do they signify that the masses have been stronger than the leaders. As a rule, they mean merely that a new leader has entered into conflict with the old, and, thanks to the support of the mass, has prevailed in the struggle, and has been able to dispossess and replace the old leader.[32] The profit for democracy of such a substitution is practically nil.

Whenever the Catholics are in a minority, they become fervent partisans of liberty. In proof of this we need merely refer to the literature issued by the Catholics during the *Kulturkampf* under the Bismarckian regime and during the struggle between Church and State which went on a few years ago in France. In just the same way the leaders of the minority within the socialist party are enthusiastic advocates of liberty. They declaim against the narrowness and the authoritative methods of the dominant group,[33] displaying in their own actions genuine democratic inclinations.[34]

As soon as the new leaders have attained their ends, as soon as they have succeeded (in the name of the injured rights of the anonymous masses) in overthrowing the odious tyranny of their predecessors and in attaining to power in their turn, we see them undergo a transformation which renders them in every respect similar to the dethroned tyrants.[35] Such metamorphoses as these are plainly recorded throughout history. In the life of monarchical states, an opposition which is headed by hereditary princes is rarely dangerous to the crown as an institution. In like manner, the opposition of the aspirants to leadership in a political party, directed against the persons or against the system of the old leaders, is seldom dangerous. The revolutionaries of today become the reactionaries of tomorrow.

Notes

1. J. K. Kochanowski, *Urzeitklänge und Wetterleuchten Geschichtlicher Gesetze in den Ereignissen der Gegenwart* (Wagner, Innsbruck, 1910) p. 10.

2. Adolf Weber, *Der Kampf zwischen Kapital und Arbeit*, (Mohr, Tübingen, 1910) p. 369.

3. Concerning the varied character and the intensity of such activities, the socialist deputy Guido Podrecca has written a charmingly humorous sketch entitled 'The Joys of a Deputy' (Le Gioie del Deputato), *Avanti!*, vol. 13, no. 44 (Rome, 1910).

4. Cf. Pio Viazzi, 'Le Gioie della Deputazione', *Rivista Populare*, vol. 15, no. 11.

5. Cf. the account given by Oddino Morgari, *Avanti!*, 12 August 1909.

6. Cf. pp. 74–5 of the original book.

7. A special chapter (pt 4, chap. 6) is devoted to this question in the original book.

8. Speech made by Ferri at Suzzara, reported in *Stampa*, vol. 47, no. 358 (27 Dec 1909).

9. Giacomo Montalto, *La Questione Sociale e il Partito Socialista* (Società Editrice Lombarda, Milan, 1895) p. 81. The description of the landed gentry as 'the rich' is a striking confirmation of the truth of Sombart's view that, in the case of a hereditarily gentle class, wealth is a natural attribute, psychologically and socially congenital, qualitative rather than quantitative. Werner Sombart, *Die deutsche Volkswirtschaft 19. Jahrhundert* (Bondi, Berlin, 1903) p. 542.

10. M. Ostrogorsky, *La Démocratie et l'Organisation des Partis politiques* (Calmann–Lévy, Paris, 1903), vol. 2, pp. 203, 206, 363.

11. 'Experience shows only too clearly that, wherever democracy is tending to degenerate, freedom of speech and the press are the first to perish.' Wilhelm Roscher, *Politik Geschichtliche Naturlehre der Monarchie, Aristokratie und Demokratie* 3rd ed. (Cotta, Stuttgart–Berlin, 1908).

12. Cf. the discussions of the congresses of the German socialist party at Munich (*Protokoll*, pp. 255 ff.) and at Dresden (*Protokoll*, pp. 158 ff.).

13. Cf. pp. 120, 127 of the original book.

14. In Frankfurt–Nordend the list proposed for the election of delegates to the congress of Nuremberg, 1908, drawn up in accordance with the express wishes of the district assemblies of the party, contained, among eleven names, those of eight officials of the labour movement (two socialist journalists, one party secretary, one secretary of trades council, one organiser, one trade-union employee, one insurance-bureau employee and one co-operative salesman) as compared with three simple wage-earners who were not dependent upon working-class organisations (Frankfurt *Volkstimme*, supplement 188, 1908).

15. In socialist and trade-union literature this aspect of the problem has often been discussed. Cf. Filippo Turati, *Il Partito Socialista e l'Attuale Momento Politico*, 3rd. ed. (Uffici della 'Critica Sociale', Milan, 1901), p. 19; Paul Kampffmeyer, 'Die Entwicklung der deutschen Gewerkschaften,' *Annalen für Soziale Politik und Gesetzgebung*, vol. 1, no. 1, pp. 114 ff.

16. Kautsky defends this claim. 'The greater the increase in our voting strength, the greater the dearth of candidates; the more remote from the great centres of economic, political and intellectual life are many constituencies with socialist majorities, the more essential does it become that the party organisations in the individual constituencies

should not possess absolute sovereignty in the choice of candidates, but that the right of selection should be vested in the party as a whole. The best way of securing this is that in the case of candidates for the diet the constituencies should secure the approval of the territorial executive or territorial congress, and in the case of candidates for the Reichstag that of the territorial executive and of the central executive. In 1876 the party congress decided the various candidatures to the Reichstag, in so far as time permitted. But in the case of a number of candidatures it was necessary that the selection should be entrusted to the electoral committee appointed by the congress. It is obvious that there are several different ways in which the party as a whole may exercise a determinative influence in the choice of candidates. Which of these ways is the most practical need not here be discussed. The point of immediate importance is to recognise, in principle, that the selection of a candidate for the Reichstag is a matter which concerns the party as a whole quite as much as it concerns the individual constituency.' The expression 'the party as a whole' is naturally to be understood as synonymous with 'the party executive'. Kautsky continues: 'Of course the choice of candidates must not be exclusively in the hands of the party executive or of a central electoral committee. The comrades in the constituencies have to shoulder most of the electoral work and it is upon them that success in the election mainly depends. It would certainly be preposterous to force upon them a candidate whom they did not want. But on the other hand the constituency must not have the right to force upon the party a parliamentary representative whom the majority in the party has serious reasons to dislike. The local organisations must choose their own candidates in the first instance. But the candidatures must always be approved by the party as a whole. . . It may, indeed, sometimes be desirable that the party, or its executive, should itself nominate the candidate. This will be the case especially in those states in which the number of safe constituencies is extremely small. Here the selection of candidates must not be left solely to the play of local influences. The party has a right to demand that in the safe constituencies those candidates shall be run whose presence in parliament is absolutely indispensable. It is owing to the unrestricted autonomy of the constituencies that in Austria such a man as Victor Adler has been excluded from the house of representatives for two parliaments in succession, and it is doubtful if he will secure a seat at the forthcoming elections. But as far as the German Empire is concerned, the number of safe constituencies is so large, that these considerations hardly apply.' Karl Kautsky, 'Wahlkreis und Partei', Neue Zeit, vol. 22, no. 28, p. 36.

17. *Fränkische Tagespost*, vol. 39, no. 191, supplement 2.

18. 'The associations of the Young Socialists are impotent vis-à-vis their leaders, lacking the force and adroitness which would enable them to avoid the arbitrary rule of these.' Max Kette, 'Die Jugendbewegung', *Neue Zeit*, vol. 28, no. 9.

19. Ludwig Börne, *Aus meinem Tagebuche*, (Reclam, Leipzig n.d.), p. 57.

20. *Denkwürdigkeiten des Fürsten Chlodwig zu Hohenlohe-Schillingsfürst*, ed. Friedrich Curtius (Deutsche Verlagsanstalt Stuttgart – Leipzig, 1907), vol. 2, p. 376.

21. Nicolas Salmeron *y* Garcia, 'L'État Espagnol et la Solidarité Catalane', *Le Courier Européen*, vol. 4, no. 23.

22. Austin Lewis, *The Rise of the American Proletarian* (Charles H. Kerr and Co., Chicago, 1907) pp. 189–90.

23. The history of the socialist party alike in Austria and in Germany affords numerous examples of minorities which were at first pugnacious and rebellious, but which have allowed themselves to be disarmed in this manner. The leaders of the opposition to the party executive at the Salzburg congress of the Austrian socialists in 1904, and also those at the Bremen congress of the German socialists held in the same year, have since then become members of the superior order of leaders and have been elected deputies to the parliaments of their respective countries. Simultaneously they have abandoned their attitude of opposition. The most typical example, however, occurred among the Dutch socialists in the spring of 1909. Here the reformist majority endeavoured to gain control of the party executive through the criticism which was levelled against the reformists by some of the particularly hardy members of the opposition. These latter, the so-called Marxist group of the *Niewe Tijd*, had their own organ, an independent and private review; now the reformist leaders of the party proposed to create a joint review, edited by the party and therefore subject to the control of the party, on condition that the Marxists should renounce the *Niewe Tijd*. This was an extremely ingenious scheme for drawing their opponents' teeth. The democratic parties in America exhibit analogous phenomena. Ostrogorsky writes: 'La machine est prête à tout faire, même a faire aux récalcitrants une place sur le tiquet?' Ostrogorsky, *Organisation de la Démocratie*, vol. 2, p. 363.

24. The reader's attention may be recalled to what has been said on pp. 103 and 104 of the original book.

25. At the Dresden congress, 1903, the German reformists found no difficulty in voting for the so-called 'Dresden resolution' which subsequently attained an international status, having been brought forward by the French Marxists at the international congress of Amsterdam, 1904, where it was solemnly reconfirmed. It is indisputable that this motion was directed against the reformists, since it expressly condemns all participation by socialists in the government. Eleven only among the 268 reformist delegates had a sense of duty and political honesty or personal rectitude sufficiently strong to induce them to vote against the resolution. Cf. the remarks on this matter of Lily Braun, *Memoiren einer Sozialistin. Kampfjahre* (Langen, Munich, 1911) p. 512; the Austrian socialist Victor Adler, whose views are most closely akin to the reformists of the German party, wrote in the Viennese *Arbeiter Zeitung*, concerning the strategem adopted by the majority of the reformists: 'Secondly, the vote means that those who are termed reformists regard the moment as unfavourable for the decisive and open declaration of their opinions and still more unfavourable for the display of the smallness of their numbers. They are indeed so few that the minority has preferred to hide itself by mingling with the majority' (from the reprint in the *Mainzer Volkszeitung* (1903) no. 225). A counterpart to this is furnished by the action of the Italian reformists at the socialist con-

gress in Rome, 1906. Here the reformists avoided a defeat only by associating themselves (although their adhesion was repudiated by the majority) with a resolution brought forward by the integralists under the leadership of Ferri, a resolution which was expressly directed against them and which contained various matters irreconcilable with the reformist theory. Thus they met the attack against them by running away. Among the Italian reformists there were not wanting some who regarded this action as inconsistent and politically dishonourable (*Resoconto Stenografico del IX Congresso Nazionale a Roma, 1906* (Mongini, Rome, 1907) pp. 275). Among those who voted against the resolution were such men as Antonio Graziadei and Alessandro Tasca di Cutò, belonging to the old aristocracy of birth, and perhaps for this reason inclined to take a more elevated view of human dignity.

26. Cf. the Paris correspondence of *Avanti!*, vol. 15, no. 16.

27. Bryce, *The American Commonwealth*, abridged ed. (Macmillan, New York, 1907) pp. 152–3.

28. Cf. a polemic article wherein Giovanni Zibordi gives an account of a visit made by Ferri to Mantua after his political volte-face. Zibordi speaks of the 'triumphal tour' of the adored leader, and deplores how Ferri and Gatti 'passed through the region of Mantua . . . amid the hurrahs of the workers who knew no better, while accompanied by the impotent disdain and grief of the socialists who saw thus installed a dangerous dictatorship, a personal dominion which is the negation of our principles and our methods'. Giovanni Zibordi, 'Quel che succede nel Mantovano', *Avanti!*, vol. 15, no. 119.

29. Hubert Lagardelle, 'Les Origines du Syndicalisme en France', *Mouvement Socialiste*, vol. 11, nos. 215–16, p. 249.

30. It is well to remind English readers that on the Continent, and especially in France and Italy, barristers play a conspicuous part in the oligarchy of socialism, corresponding with that which in England they play in the old political parties. (Translators' note.)

31. Rinaldo Rigola, the socialist secretary of the Italian General Confederation of Labour, describes the socialist party as an oligarchy, and therefore contests its right to present candidatures for the elections and to decide the policy of the proletariat. In his view, these functions should rather be allotted to the labour organisations, whose membership is far more extensive and which could constitute themselves into a Labour Party. Rinaldo Rigola, 'Discutendo di un Partito del Lavoro', *Avanti!*, vol. 14, no. 172.

32. Richard Calwer, in a declaration to the socialist press, gives the following account of his dethronement as a party leader: '*Vorwärts* and the *Leipziger Volkszeitung* accept as a matter of principle the resolution of the party conference of the third Reichstag constituency of Brunswick, by which it was decided to repudiate my candidature in future. They do this without reflecting upon the moral poverty which the decision exhibits for the party. The dissatisfaction of the comrades in the constituency with my economic views is supposed to have increased gradually, and at length to have become overwhelming. It is strange that during the entire sixteen years during which I have been a candidate in this constituency there was not until about a year ago the slightest manifestation of dissatisfaction among the comrades in the electorate. Yet never throughout this period have I made any secret of my views. The local comrades have been familiar with them from the first and have never, for this reason, wished to remove their confidence. The alleged divergencies "in matters of principle" date from no more than a year back, having begun precisely at the moment when comrade Antrick came to Brunswick as secretary. What reasons there were to induce this comrade to attack me "on principle" I do not know. In any case, I neither had nor have inclination or time to trouble myself about personal quarrels and to dispute with comrade Antrick. *Volksstimme*, 15 Aug 1907.

33. Cf., for example, the pamphlet issued by the displaced members of the staff of *Vorwärts*, *Der Vorwärtskonflikt: Gesammelte Aktenstücke* (Birk, Munich 1905), in which we read: 'We are not here concerned merely with the moral position of the journalists within the party; the present conflict is a matter of decisive importance to the internal well-being of the German labour movement. The question at issue is that of the dignity of all the responsible persons in the confidence of the democracy. What has to be decided is whether a system of absolute publicity is to be replaced by a secret method of jurisdiction; whether open discussion is to yield to the crafty dissemination of suspicions; whether obscure intrigue is to oust comradely confidence; whether blind caprice is to be more effective than reasoned conviction; whether arbitrary opinion is to be more influential than established fact or whether, in a word, a regime of glib demagogy, of personal ambition, and the most unscrupulous place-hunting, is to be established in the German democracy.'

34. Cf. p. 18 of the original book.

35. 'When he has the power in his own hands, he ignores the laws which were made for his restraint.' Giambattista Casti, *Gli Animali Parlanti, Poema* (Tip. Vanelli e Comp., Lugano, 1824), vol. 1, p. 30.

THE POLITICS OF RECREATION

R. Frankenberg

THE interest aroused within the village by the parish council and its affairs does not extend to what the parish councillors do outside the village. Villagers are indifferent as to whether parish councillors attend the meetings of the County Association or not. Although the parish council acts from time to time for the village as a whole and negotiates on behalf of the village with outside powers, it is not a symbol of village unity and cohesion against the outside world.

During 1953 the village football club provided such a symbol but this was not, of course, its only function in village life. The fact that the outside prestige of the village would be judged by the football club's public appearances did, however, have the effect of giving football a central place in village social activity. In the village details of the internal struggles of the football, flower show and fête committees were of paramount importance, trivial though they may seem from outside. This is generally true of village affairs, as is demonstrated if one attempts to read a local newspaper from an area with which one is not familiar; its repetition of detail soon palls on the outsider. Furthermore, internal village events provide a private time scale for villagers which a visitor soon adopts. When, on a brief visit from the field, I read a paper to the Manchester Seminar, my colleagues laughed at my dating of events. I used such phrases as: 'Just before the Carnival', 'after the Coronation Tea' and 'before the Sheep-dog Trials'. To me, still immersed in Pentrediwaith, the joke had to be explained.

Football is only one social activity which brings the village into contact with other villages, but in 1953 it happened to be the most important. There were, even then, others. Eisteddfod 'parties' of singers and reciters attended local Eisteddfodau, drama groups performed in other villages. In other years there was a choir and a brass band which travelled about the countryside to give performances. Such trips enlarged the villagers' acquaintanceships and emphasised to other villages the existence and prowess of Pentre people.

Football brings other village teams in, as well as sending Pentre villagers out. This bringing-in is also achieved by the sheep-dog trials which have been held annually for over fifty years, by the flower show, and in 1953 by the carnival. Sectional groups within the village also brought visitors in. There was a performance at the Baptist Chapel of *The Messiah* by a choir from outside and debating teams were exchanged by the Baptist Young People's societies of Pentre and other villages. All these activities, however, which at different times have made Pentrediwaith, or groups of its villagers, known and praised outside the valley, were overshadowed in 1953 by the activities of Pentrediwaith Amateur Football Club and associated events in the village.

Association football is the established sport of this part of North Wales. I cannot say how it came to be so, for this is a problem soluble only by historical research. But once established, it is possible to see some of its advantages over other possible recreations. It is a sport, in contrast to sheep-dog trials and slate-splitting competitions, which has a more than specialist appeal.[1] The equipment needed for football is relatively inexpensive and the preparation for a game or a competition is not very arduous. The season is long, from August to May, but leaves the summer free for gardening and holiday activities. Football is very nearly an all-weather game. The Welsh Football Association and the Welsh National League provide a framework in which a village team can take its place. Since this framework exists and other villages have teams it is considered that each important village should be represented in its local League in order to maintain its prestige. The national organisation takes (or should take) much of the burden off local organisations since it fixes the length of the season, arranges the fixtures and lays down the rules. The League provides referees and specifies kick-off times. The village has merely to muster eleven players and some officials and provide a pitch. The rest is provided from above. Finally, association football is above all the British national game. Whatever local and class allegiances there may be to rugby union or rugby league, league and county cricket or to tennis and still more localised games like marbles, bowls, darts or archery, association football is the one sporting interest and topic of conversation which can cut across all class and regional boundaries.

Pentrediwaith's football club provides participation in this national game for both spectators and players. The spectators keep their eye in for North Wales Coast matches and the Welsh Amateur Cup, for Third Division (North) League matches and cup-ties and for Internationals and the First Division farther afield. To have a village team, especially when composed of local lads, fosters the fanatical partisanship which gives savour even to professional League football in the towns and gives Pentre football an intense importance in the village. The honour of the village and its place in the outside world are at stake in each game and in the day-to-day conduct of the club.

The players enjoy the game, the exercise and the local prestige. They sometimes hope to advance to Oldham or Wrexham in the Third Division or to the better North Wales clubs. One or two

Pentrediwaith youths in recent years have had trials for Manchester City. It is not forgotten in Pentrediwaith that the fabulous Billy Meredith, Welsh International and player for both Manchester City and United, came from a village in North Wales.

The fact that everyone in the village, almost without exception, is pleased when Pentrediwaith distinguishes itself on the football field does not mean that the village unites, in amity and co-operation, to make the village football team a success. The very intensity of interest that the game arouses leads also to divergencies of opinion. Committees split into overtly and latently hostile groups, often along lines predetermined by family dislikes or by insults incurred in entirely different situations but carried over into the organisation of recreation. The divergent groups are not congruent with those already discussed in political organisation, but they overlap and cut across them. Church and Chapel members are, for example, to be found on both sides of a football controversy.

The Chapels do not actively oppose football although the more austere deacons and ministers may regard it as a frivolous occupation. Many of the activities which were associated with the raising of funds for the football club were very much opposed by the Baptist minister in Pentrediwaith. His opposition to dances, whist-drives and raffles did not, however, seem to discourage his members from participation in these events, or even from taking an active part in their organisation. The Anglican vicar was an active supporter of the club and a vice-president; he contributed generously to its finances and, whenever he was free to do so, attended its matches. Considerations of Church and Chapel did not enter to any important extent into the deliberations of the committee or its initial election.

Thus the 1952–3 football committee had a Baptist chairman and treasurer and a Church secretary; five of the rest of the committee were Church and seven were Chapel. The Supporters' Club Committee, in the same season, had a Church chairman and secretary and a Baptist treasurer; six of the rest were Church and seven were Chapel. In the football committee on which I served in season 1953–4 the chairman, secretary and treasurer were all Chapel members. There were two English outsiders, two other Church and three other Chapel members among the regular attenders. The active supporters' committee members in the 1953–4 season had a Church chairman and secretary and a Baptist treasurer; six of the ordinary members were Chapel members and one was Church.

Both Church and Chapel members were equally shocked when a sweepstake was suggested. The only difference in behaviour which arose from denominational differences was that Chapel members were more likely to miss meetings because of the demands of other activities organised by the Chapel. Generally speaking, membership of a particular denomination did not seem to me directly significant in determining behaviour within the football and supporters' committees or in other activities connected with football. When differences had arisen along other lines, however, denomination was sometimes used, in private discussion, as the last straw added to an already heavy indictment: 'Those Church people are always the same.' The effect of denominational differences was then to help in fixing individuals in their social positions in the village by adding another consideration to the individual's unstated problem: 'Can I co-operate with X in this matter?'

The social divisions which are significant to the organisation of football are apparent when the structure of the football and supporters' clubs is examined and when the sequence of events during 1953 is described. There are in Pentrediwaith two clubs whose primary concern is, or was until mid-1953, the carrying on of football. They are the Football Club and the Football Supporters' Club. We shall see that the clash of interested groups within the individual clubs is mitigated by the fact that after a serious dispute the defeated group withdraws from the committee and continues its criticism from outside. Some members of these defeated groups do not stop at criticism but actively work against the interests of the club they have left. Thus, during 1953, notices of forthcoming matches and the chosen team were mysteriously torn from the tree where they had been fixed. Other such examples will appear in the following pages. The withdrawal of support for a social activity by groups of villagers eventually results in the death of that social activity. The football club reached this pass in 1953–4 and did not survive into the next season. From what villagers told me it appears that this fate also overtook the brass band and the male voice choir, but I have not documentary historical material to confirm this.

Whatever the internal divisions in the football and in the supporters' club, the major division in this context is between the *football* and the *supporters* – that is, between the men and the women. This was clear not only in the progress of the carnival–football dispute, and in the proceedings of the carnival committee itself, but also in the daily behaviour and conversation of villagers. For example, when particular men failed to attend whist drives organised by the supporters, they were criticised very severely by the women. 'Where would the football club be without the women to make money for it?' members of the football committee were asked. The usual reply, not often given directly, was to ask what the women would find to do if they had no football club to support. On one occasion a man who was very active in the football club attended a supporters' whist drive after threats by the women to withdraw support from the football club if he did not. He took first prize. He regarded this as 'one in the eye' for the supporters and told the story to other men with great mutual enjoyment. The whole incident only served to intensify the resentment of the women.

The argument about whist-drive attendance was constantly repeated between different men and women. It is of course true that in actual fact many men enjoyed the whist drives. Furthermore, some women not only came to watch home matches but even travelled with the team to away games. Although such people were not regarded as abnormal and women supporters were expected to attend at least home matches, villagers behaved as if football itself was an activity which was of interest and benefit only to men. On the other hand social activities like dances and whist drives, organised to raise funds for the football club, were spoken of as a chore that the women carried out for the benefit of the men. Both these views were put forward in the first place by the women, but they had secured their acceptance, at least in public, by the men. In fact, as I have suggested and will demonstrate, if football had not existed to provide a reason for organising events, some other excuse would have been found.[2]

It is necessary to make a brief digression here to describe the very strong sense of obligations of reciprocity that pervades Pentrediwaith life. The relations of the supporters' club and the football club were coloured by this attitude to mutual obligations. The supporters' club gave £70 to the football club in season 1951–2, £50 in season 1952–3, and £60 in 1953–4. But these gifts were not free of obligation. In Pentrediwaith it is not considered comfortable to receive a free gift. A gift creates an obligation, which must be met. This principle extends from the slightest transaction between individuals to the dealings of corporate groups. To a Pentre villager getting something for nothing from another villager is not an achievement, it is a disgrace. Thus when a woman had taken morning tea with her sister daily for two or three weeks she felt it necessary to bring a little tea in a paper bag as recompense. A young man invited to watch the Hungary–England football match on television was very doubtful about accepting the invitation lest he incurred thereby an obligation that he had not the means to repay. This scrupulousness does not always extend to outsiders. When a publican installed a television set in his bar he found it increased attendance but reduced his takings; even regular drinkers found the television so absorbing that they forgot their obligation to drink as well as watch. Although 'co-operation' on the scale described by Rees and Huw Evans [3] is not carried out in Pentre, some villagers have arrangements to help local farmers at critical times in the farm year. They are rewarded for these services by their meals and in kind. For example, one family who keep chickens are given corn in return for their harvest services. Those who help with potato-lifting receive sacks of potatoes. When I coached a farmer's son in English, I did not of course receive payment for my services, but I seldom left the farm without supper and never without a rabbit, a dozen eggs, a bag of potatoes or some similar gift.

The wife of a villager who did local building jobs shared out her food registrations (under rationing) between two of the four village grocers and bought bread from a third. She had milk delivered by both village milkmen and bought additional bread from the baker. Her husband in return received some of each tradesman's building work which they shared out in the same way. When in 1953 advertising space on the carnival programme was sold, one woman shopkeeper misunderstood and did not buy any. When the programme appeared and her name was not mentioned, she not only refused to attend the carnival or help it in any way, but also tore up a copy of the programme offered to her by a seller on the village's main street. Despite Chapel rivalry, members of different Chapels and the Church support each

other's functions, and no door-to-door collectors in the village are turned away empty-handed.

Reciprocity has helped the village to keep going in times of difficulty, and has united the village and its participant groups into a coherent whole; but when it is not observed, or is believed not to have been observed, reciprocity can divide as well as unite. Thus, in the relations of the football and supporters' clubs reciprocity plays an important part. They are united by a common interest in keeping the football club going: when this fails, as it first began to do in 1952, breaches of reciprocal obligation widen the gap.

In theory the whole village are members of both clubs. Both clubs have, or should have, annual general meetings which all villagers are entitled to attend. In fact very few villagers outside the committees do attend the public meetings of either; when a villager refers to the supporters' club or the football club he usually means the committee. But the committee are regarded as representing the whole village; like the parish council, they are subjected to constant informal criticism which, as in the case of the parish council, is made irritatingly more effective by its refusal to come into the open at public meetings.

Pentre people have other infallible sanctions to keep these committees under control. If villagers stay away from events organised by the committee, they gain the double satisfaction, first, of placing the committee in an embarrassing position financially and, secondly, of being able to blame it for the plight into which it has in fact been forced by public action. Villagers who first oppose and then kill an event in this way can and do say: 'We told you so!' Thus although only a very small number of villagers are active in the day-to-day running of the football club, these few receive informal advice and criticism without stint from the majority. The duties of the committee members, although light if shared by many, can become very arduous, and Saturday-night criticism is not the least of the sacrifices of personal freedom and other discomforts involved in accepting office. There is little open competition for places on the committees (although there may be a struggle to displace officers). Members once elected rarely resign during the year. This is because committee members themselves accept the values of the community and fear both the downfall of the club and the blame and loss of face the committee members would suffer if they allowed football to die out. At the time of the annual general meeting, however, they do succeed in finding excuses to leave. Different officers have led the club each year in the three seasons, 1951–2, 1952–3 and 1953–4. Those of 1953–4 declined to carry on another year, and the club no longer functions.

In some committees there is an almost irresistible social pressure against resigning. One woman told me how she became ill and wanted to give up her village social commitments. When she announced that she was leaving the British Legion (women's section) committee the secretary became very cold in her attitude towards her, and the would-be resigner stayed on to prevent ill-feeling. She also stayed on the supporters' committee because the secretary said he would resign if she did. Then other members said they would not stay if the secretary went. By this time the woman felt that public opinion was against her, someone said she was trying wantonly to destroy the whole club, and she gave in. Not until three years later, when her ill-health was apparent, did she succeed in resigning. Her case was not untypical.

THE FOOTBALL CLUB

The football committee met once or twice a week to choose the team for the following Saturday and to transact any other business as it arose. The other business often concerned disputes with the League over the non-fulfilment of obligations on both sides. Committee members were expected to travel, at their own expense, on the bus which carried the team to away matches on alternate Saturdays. One of the committee acted as linesman and another as trainer. In addition, members of the committee had to play occasionally as substitutes for players who had failed to turn up. For home games the field had to be marked with lime, the grass cut occasionally and thistles uprooted.

These last three jobs were all arduous and unpleasant. Usually the lime was applied by hand in powdered form. One just filled a bucket with the powder and sprinkled it in handfuls along the lines (or what seemed to be the lines). This job ruined one's clothes, dried one's throat and stung one's hands. It was very expensive in lime and on windy days the marker appeared to have been rolling in flour. Later in the 1953–4 season the club acquired a mechanical

marker which employed a lime solution, but the condition of the field reduced its efficiency and made it difficult to use. When there was snow on the ground sawdust was used for marking. Other preparations on the field, which was a pasture during the week, were also arduous and unpleasant. The grass was cut with scythes or even sickles and mole hills had to be flattened. Occasionally turf displaced by pigs had to be replaced. The players had to take their chance of avoiding the cowpats as the committee had no inclination to remove them.

The nets had to be placed in position before home games and the balls inflated and fetched to the pitch. Two committee members had to collect the money at the gate, while another fetched tea from the village at half-time and paid the referee's fee and expenses. After the game the nets had to be taken down again. This was a wet, cold, dirty and awkward job because the nets were tied to the cross-bar with creosoted string and there was no step-ladder. Someone else had to collect the players' dirty clothes and see that they were washed and in good condition for the following week. Finally, according to Football Association rules, the committee is also responsible for the safety of the referee; this is not always a light burden. Thus the committee members had chores to perform as well as decisions to make, and some of them, like marking the field, were both time-consuming and tedious. These jobs were the prerogative of committee members only and the suggestion at committee meetings that others be 'allowed' [*sic*] to help was at first strongly opposed, although it was eventually adopted.

As the Welsh language served to determine the class composition of the parish council, so the existence of chores limited to some extent the economic class status of members of the football committee. Excluding myself, all those who served as ordinary committee members in seasons 1951–2, 1952–3 and 1953–4 were working men. Even the officers were not far removed. The chairmen during the first two seasons were in fact employers of labour in a very small way, but they worked themselves side by side with their employees. The vice-chairman, the manager of the local Co-operative Stores, helped each week to take the nets down and to collect money at the gate, but he was implicitly excused from the various tasks of preparation of the field. In season 1953–4, the chairman, secretary and treasurer were all white-collar workers, and on several occasions one of them complained to me that cutting grass was not the sort of work he was accustomed to; it made him feel, he said, 'like a bloody navvy'.

Although the committee had, in theory, a hierarchic structure, its leaders lacked authority. Decisions had to be taken in accordance with 'the principle of unanimity'. Consequently necessary tasks were frequently not performed. On one occasion only seven players turned up to an away match and the linesman, trainer and other committee members played. Once the pitch was not marked at all and the referee refused to start until at least some of the salient points had been indicated. Home matches almost invariably started late and the club usually had to pay a fine at the rate of sixpence per minute to the League. Sometimes the half-time tea did not materialise or the referee remained unpaid. I do not remember a single Saturday when everything went off without a hitch.

Apart from their direct responsibilities to carry out the tasks necessary to run the football club, committee members had the incidental duties of attending social functions organised by the supporters' club in the village and of buying the raffle tickets which members of the supporters' club take round at every home match. Alleged failure to fulfil these obligations gave rise to continual controversy.

The committee which carried out all these functions had as the senior official in its formal structure a president. He was expected to give a handsome donation to the club but not to preside. In 1952–3 the president was the Welsh-speaking publican of one of the three village pubs. It appears that he did actually preside over informal inquests held in his bar. He gave a £5 donation.[4] In 1953–4 he declined to continue in office, and his place was taken by the English-speaking proprietor of the principal village grocer's shop. He was a newcomer to the village and he did on one occasion preside at a committee meeting but, trained in English public school 'rugger', he apparently did not feel at home and did not come again. He remained a generous supporter of the club.

Next to the president were ten or eleven vice-presidents – local gentry, shopkeepers and publicans who owed their position to their ability and willingness to help financially. Below the vice-presidents was the club committee proper with a chairman, a vice-chairman (representing the Women's Supporters' Club of

which, though a man, he was chairman), a secretary, a treasurer and ordinary committee members. Most committee members in season 1953–4 were ex-players for the club. Committee members who still play are not encouraged because of the obvious difficulty of discussing their performances. One man who was very keen on the football club refused an invitation to serve on the committee on the grounds that it would prejudice his chances of being chosen to play for the club. The team captain was an ex-officio committee member but never attended. Only three members of the committee had never played for the club: myself, a Londoner who had come to live in the village and worked in a nearby colliery and the 'supporters' representative'.

The membership of the football committee was exceptional at the beginning of season 1953–4 in that there were five coal-miners serving on it. This was probably the result of disputes in season 1952–3 between the miner secretary and other committee members. It was, however, very characteristic of Pentrediwaith committees (other than those concerned with local government) that their membership did not consist of isolated individuals but of groups united by kinship or other ties. These groups voted, and from time to time resigned, as units. The miners' group in the 1953–4 committee was one example of this; in the previous season an informal drinking-group led by the chairman was another. This latter group resigned as a body at the end of season 1952–3, but when things were going badly in 1953–4 offered to come back as a group. Their offer was accepted, but they changed their minds before the next meeting. When the miner secretary (12) resigned in 1953 he resigned alone, but he continued to discuss the affairs of the club and to exert an influence on them through his informal relations with the miners who remained. One by one during the season they also resigned or just ceased to attend.

The new secretary, Humphrey Biggs (6), also collected a group of this kind by pressing his father and brothers into service during a period when it was difficult to find people to do the necessary jobs. Long before this, however, villagers were blaming the football club's lack of success on the fact that it had got into the control of 'those Biggs'. (The numbers in brackets after names and officials refer to Fig. 28.1 on p. 545.)

There was an opposing group within the committee, but conflicts actually within the football club never became very intense. The crises within the club were acute but not chronic, since they usually led to a conflicting group's resigning or becoming inactive. Consequently, as we have seen, by the end of 1953 there was scarcely anyone left prepared to serve on the football committee. Even the last of the eligible strangers, that is, those sufficiently humble to do manual work, had departed. As I have suggested, this seems a characteristic development.

The changes in the committee during 1953–4 demonstrate this process of reducing the numbers available to serve, but they are linked also to changes in the other, and perhaps most important, section of the club – the players. The tendency in the three seasons under review was for the players to be drawn less and less from Pentrediwaith itself. In some games in 1953 there was only one villager in the team, yet I was told that when the game was first revived after the war the team consisted entirely of Pentre players. The first post-war team included three pairs of brothers all from this village. The twenty-six registered players in season 1952–3 included thirteen from the village. No records were kept of which of these actually played, but from my own knowledge of the players and from discussions with villagers I can say that five of the Pentre players on the list would be chosen only if it were unavoidable owing to the absence of other outside players. While I was in Pentrediwaith I kept a weekly record of the gate-money and raffle money, who was chosen and who played, which I present in table 28.1.

The post-war founders of the club remained in office as chairman and treasurer for five years until the end of the 1951–2 season. The secretary at that time was a young bank clerk who has since emigrated to Australia. By the end of the 1951–2 season the club was well established, and the question first arose which has bedevilled its activities ever since. The committee were divided as to the aims of the club. Should it be just to provide football for local boys to play and local men to watch, or should it go all out for the championship of its division in the Welsh National League? A necessary corollary of the latter decision was bringing outsiders into the team, as by 1952 it was already clear that local players on their own could not reach the required standard.[5] Decision was reached in favour of all-out competition in the League and the chairman, Tom, and treasurer, Philip (1), champions of the opposing viewpoint, retired from the club.

TABLE 28.1

Week	*Gate* £ s d	*Raffle* £ s d	*Number of villagers in team* *Chosen*	*Played*	*Result*	
1	4 6 0	1 4 8	5	8	Lost	9—1
2	AWAY		8	8	Lost	2—1
3	5 0 0	1 15 0	3	3	Lost	7—2
4	AWAY		3	5	Lost	7—1
5	CARNIVAL					
6	3 0 0	1 13 0	2	3	Lost	7—1
7	3 0 1	1 10 6	3	3	*Won	4—3
8	AWAY		3	4	Lost	5—2
9	AWAY		3	4	Lost	11—2
10	AWAY		3	3	Won	5—3
11	2 10 0	19 0	3	3	Lost	7—3
12	AWAY		2	2	Lost	6—0
13	1 19 8	10 0	3	3	Lost	5—3
14	2 11 0	13 6	2	2	Lost	3—0
15	AWAY		1	1	Lost	8—4
16	1 5 0	No raffle	1	2	Lost	4—2

*Disallowed.

I was not, of course, in Pentrediwaith at this time, but from conversation with villagers I have formed some impression of these early post-war years. The team was at first not only local but also strong, but its members were near the end of their playing careers. Some of these players were on the committee in 1953 but had retired from active playing. The young men in the original team have left the district to work elsewhere. Unlike season 1953–4, the early years were years of success, but the committee was nevertheless not always at peace. There were even then resignations and quarrels, the fruits of some of which were still apparent in 1953.

At the end of 1951–2 the team stood third in its League Division. They had up to that time played on a field adjoining the Castell road, between the Council school and the minister's house. This had the disadvantage that it was easy to watch the game from the outside without going into the field at all. In this way villagers could avoid both the entrance payment and getting their feet wet. The danger to the club's revenues arising from this situation was partly met by having a collection among the non-paying spectators, and the response to this expedient was improved by having a policeman to make the collection in his helmet. The new committee, with Charley (12) as secretary, moved from this field to a more remote pitch a little way outside the village and surrounded by other fields. On this new site only paying spectators could see.

This move had its disadvantages. At the old field watching football on Saturday was often a casual affair. People on their way to and from the village had to pass the field. Sometimes they stopped for a moment to watch, became interested and stayed. The general uproar during an exciting game could be heard from all parts of the village and attracted people to the ground. This gave rise to an amusing incident during one particularly hard-fought game with a team regarded as a traditional rival. On this Saturday afternoon a villager going off to shoot rabbits was on his way up the hill leading to the church when he heard a particularly loud roar from the football ground. He returned to the village to see what was going on, became more interested in watching the game than continuing his shoot and remained. The referee made decisions which the visitors did not approve of and they later complained to the Welsh F.A. At the inquiry which followed, the visitors' officials insisted that not only had the Pentrediwaith crowd been aggressive and threatening but that one of them had even brought a shotgun to intimidate the referee.

At the field used in 1953, to come and watch the game was always a deliberate act. No one passed the field or visited it except for the specific purpose of watching football. Some villagers felt, I believe with reason, that by moving in this way outside the village physically, the football club was also isolating itself to some extent from the social life of the village. Only real enthusiasts would follow a losing team in such circumstances. In my opinion, it is highly probable that this was a factor in producing the falling off of feminine interest in the game, and all that that implied in the loss of financial support. For the women of Pentre were often busy on Saturday, as on any other afternoon. For them the moving of the field from a central point in the village to right outside was almost equivalent to moving it to Castell or Tonmawr. The move made football virtually inaccessible. Many women lost interest also when the team came to be predominantly composed of outsiders. The combination of a 'foreign' team and a distant field adopted by the committee at the outset of season 1952–3 brought into the open the already existing conflict of interests between the

men of the football club and the women of the supporters' club. As I have suggested, in Pentrediwaith, particularly now that most men go away to work, the full support of the women is crucial for the success of any activity.

At this critical point in the external relations of the football club a new secretary, Charley (12), a coal-miner, took office. This internal change had significance beyond a mere change of personalities. Previously secretary, chairman and treasurer worked together with the committee to choose the team. The secretary had the responsibility of notifying the team and therefore had no vote in committee on the selection or placing of the team. I often heard him in 1953 arguing with players and pointing out that he was not responsible for changes in the team, but that the committee was. This worked in practice as well as in theory as long as the team was local. The secretary at the outset had no more power than other committee members, and perhaps had less. Furthermore, he was a player himself and a young man, whereas Tom, the chairman at this time, was much older and used to authority. Tom was Welsh-speaking, Baptist and a builder who worked beside, as well as gave orders to, his men. He was sufficiently outside and above the group he led to exert authority over them, but sufficiently similar in outlook to be listened to, and worked with, by the other committee members in close co-operation.

When the principle of a team of outsiders was accepted, the secretary's position changed. The village club had to compete with other local teams for the services of even local players. Competition for players from farther afield was still more intense. It usually came about, therefore, that players brought in from other districts were in fact either workmates or friends of workmates of the secretary. The relative position of the secretary and the rest of the committee was now reversed. The secretary no longer merely notified the players chosen by the committee; he now nominated players to the committee, secured their selection and was the committee's only means of communication with them. If a secretary says that X will not play inside-left, and X is a village boy, the statement is easily checked. But if X comes from a relatively distant town and works at the next laboratory bench to the secretary, he is difficult to get hold of for questioning. The secretary was in fact invested with power, and this inevitably led him into conflict with the committee.

I know that this situation existed in season 1952–3 but I cannot give detailed examples of it. In season 1953–4 there were many examples. One player from outside whom the committee wished to play half-back was said to have refused to play anything but inside-forward. Only when the secretary was ill one week was contact made with him by post and his position changed. On another occasion it was decided not to drop a player who was getting old and replace him by another. Both players involved were outsiders from the same works. The change was made, despite the committee, 'at the older player's wish'. The committee had no opportunity to check this in time.

The officers for the 1952–3 season, apart from the secretary, resigned at the end of the season on the ostensible grounds of other more pressing work. They said in private it was because the secretary was too domineering and chose his own team. In season 1953–4 one group resigned from the committee for the same reason after an open discussion in committee.

Pentre villagers participate in village activities not as individuals but in small informal groups composed of from three to six individuals. These groups join, participate and resign from committees *as* groups. In this context this group behaviour has two results: firstly, the secretary collects his own group of supporters in the committee, as Charley did with the miners; secondly, the secretary's power is increased by the fact that the players also participate in informal groups among themselves.[6] Given that the players come from the secretary's place of work and that they are comparatively unknown to the rest of the committee, he is able to tell the committee that if Y is dropped from the team as the committee wishes, X and Z will not play because X, Y and Z always spend Saturday afternoons together. Further, since such groups overlap, if Z does not play, A will not, and Pentre may lose B on A's account. At one point in season 1953–4, two players from a First Division team of the Welsh National League transferred to Pentrediwaith. The Pentrediwaith committee members were naturally curious as to why they wished to leave a winning First Division team for a losing Second Division one. The secretary replied that they did so to be with two of their workmates and friends already playing for Pentrediwaith. On another occasion two Trefawr brothers who had previously played for Pentre were going to return to Pentre because they objected to the dropping

by Trefawr of their friend, the Trefawr centre-half. Unfortunately for Pentre the proposed substitute centre-half broke his arm and could not play, so the three friends stayed in First Division Trefawr.

Thus whether he abused his power (as committee members alleged) or not, the secretary had power, once the principle of the outside team was accepted, to dominate team selection. This brought him into conflict with the rest of the committee. During this conflict he might, and did in Pentrediwaith in both 1952–3 and 1953–4, develop his own group of adherents. The chairman, if he were effective, would be outside both the secretary's group and their opponents. The role of the 'stranger' will appear in the detailed discussion of the events of season 1953–4.

When the founders left the football club in 1952 and Charley (12) became its secretary, there were fifteen committee members. The chairman led one group which I shall call the Green group: it had ten members. In addition there were three older men, a miner friend of the secretary, and Humphrey Biggs (6), whose father was one of the Green group. Despite their resignations six of the Green group were elected on to the committee again in season 1953–4. Not one of them attended a single meeting. Nor did they attend the general meeting which opened the 1953–4 season. At this meeting Charley (12) was re-elected secretary, and Philip (1), one of the founders, was recalled to the committee by being elected chairman in his absence. Five miners, (7, 8, 9, 10, 11), friends of Charley, one of them an Englishman (7), were elected on to the general committee, and Humphrey Biggs (6) was re-elected as well. Timothy (3), the brother of the original secretary, was elected treasurer. He played for the club and was a bank clerk in Bigtown. Percy (2), the manager of the Co-operative Stores and chairman of the supporters' club, was elected to represent the latter. Arnold (5), Percy's assistant in the stores and an ex-player, was also elected. Finally I myself (4) was elected to the committee on the stated grounds that I always attended matches and had been keen enough to attend the annual general meeting. I pointed out that I attended out of duty, but this was not considered a bar to my election if I were willing, which of course I was.

This then was the committee which started the season 1953–4. The decline in local support and in the local composition of the team which had started in 1952 was reaching its climax and the season began, as it was to go on, in conflict.

The chairman, Philip (1), was no exception to the rule of stranger chairmen of committees, but even he did not remain active long. For one excuse or another he stayed away from the committee meetings; he ended by breaking with the club on the grounds that he had too many enemies on the committee. Philip was Welsh-speaking and the clerk at the village branch of a national bank (a two-man establishment). He attended the Baptist Chapel but had no home in the village. He lived at the hotel and often went home for weekends to his own village. He was previously the club goalkeeper, and had been treasurer when the club was revived after the war. Percy (2), the vice-chairman, who presided more often than the chairman, was, as has been mentioned, the manager of the 'Co-op' and an English-speaking Welshman born in a nearby village. In the absence of both these men I was usually asked to take the chair.

As we have seen, the founder-chairman, Tom, was not an outsider, as (although he was born in a neighbouring parish) he had many relations in the village, spoke Welsh, attended the Baptist Chapel and was a parish councillor. He was nevertheless set a little apart, as he had never worked in the quarry and was in 1953 an employer of labour himself on a very small scale. Until 1926 he ran a family coal merchant's business which he had inherited. The second chairman, Green, had a similar background and ran his own small business. It was not until crisis was reached in 1953 that strangers were drafted on to the football committee in force. Indeed, the first comments of villagers when the new committee was announced were to the effect that it seemed that village football was to be run by 'foreigners' that year.

In the interim period between seasons 1952–3 and 1953–4 came Coronation Week, and the football club obtained permission from the Welsh F.A. to hold a knock-out competition out of season as part of the local festivities. The Football Association imposed the condition that the local Coronation organisers must approve the football club's plans. After an exchange of letters the Coronation committee did approve. A characteristically reciprocal agreement was entered into. The football club agreed that the profits, not then expected to be large, should be devoted to the general Coronation fund or the losses met by it. In return the

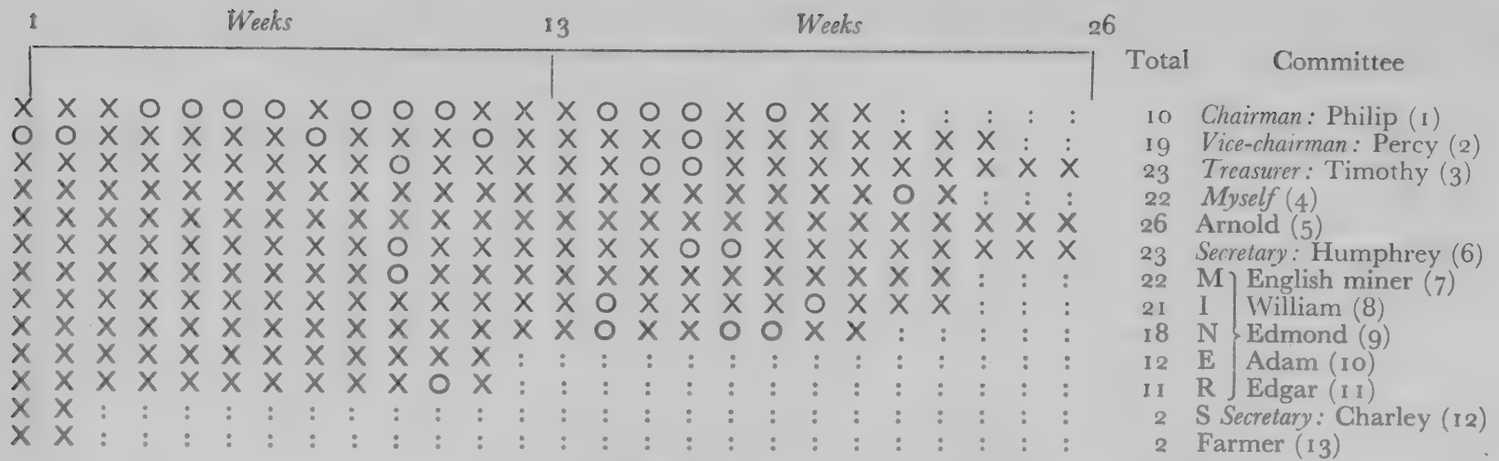


Fig. 28.1 Football committee members' attendance record
(Six of previous year's committee re-elected but did not attend.)

William (8) and Edgar (11) are brothers. A third brother was the team captain and Humphrey's (6) best friend. Percy (2) and Arnold (5) work together as manager and assistant in the local 'Co-op'. Edmond (9) has a brother, Edward – numbered (25) in Figs. 28.2 and 28.3. Edward (25) played in the team until his friend Charley (12) resigned. Percy (2) is also Chairman of the Supporters. Philip (1) and Timothy (3) are bank clerks. Humphrey (6) also has an office job.

X denotes attendance, O denotes absence, from a particular meeting.

parish council, who were controlling the Coronation committee, would provide help during the matches at the gate and in other necessary tasks. The competition was successfully held. Matches were played each evening of Coronation Week, and the final tie played off on the Saturday after the Coronation itself. It was in fact the only successful part of the celebrations in that week. The rest were ruined by the weather.

The Coronation cup final was won by one of the two Pentrediwaith teams entered, but only after some unfortunate incidents which had the immediate sequel that the other finalists left the field and the village some twenty minutes before the end of the game. The ultimate sequel was that the visiting team were rebuked after a local F.A. inquiry and a player was suspended. The following paraphrased report from a local newspaper gives the facts as stated by both sides. The visiting team's allegation that Pentre broke the rules of the competition is, I believe, true, and they had some grounds for complaint if no excuse for their own behaviour.

TEAM WALKS OFF IN CORONATION FOOTBALL FINAL

There was a sensational ending to the final of the Pentrediwaith football knock-out competition on Saturday, when Pentrediwaith 'A' met a Shropshire team. Twenty minutes before full-time, and after an incident in the goal mouth, the visiting side left the field.

It was stated on behalf of the home team that twenty minutes from time, when Pentrediwaith were leading 3–2, the referee awarded a penalty against their opponents for an infringement in the penalty area and at the same time ordered the opposing goalkeeper off the field for – allegedly – striking a Pentrediwaith forward. It is alleged that the visiting team left the field rather than accept the referee's decision.

Their secretary alleges that Pentrediwaith 'A' included in their team four players who had played earlier in the week against his team for the Pentre 'B' team in the same competition. He alleges that when his side strongly objected, they were told that if the four players were not allowed to play there would

be no match. Rather than disappoint the large crowd present, which included a fair representation of their own supporters, they agreed to play.

After about 22 minutes in the second half, the secretary says, the goalkeeper was fouled and the referee sounded his whistle for a free kick. The goalkeeper, however, got hold of an opposing forward, whereupon the referee appeared to order the goalkeeper off the field. There was some delay and the referee then indicated a penalty. There was some argument and further delay, and quite a number of spectators were on the field of play when the visiting team left the field, says the secretary.

While he says that his team should not have left the field, he states he is submitting a report of the incidents to the Welsh F.A. and to the local branch of the F.A. with a copy to Pentrediwaith. His complaint is against the organisers of the competition.

Pentre villagers were unanimous in blaming the visitors for what had occurred, and hostility between the villages still survives. On one occasion later in 1953 a group of Pentrediwaith young men were driven from a public house in Bigtown by youths from the other village.

I did not count the number of people present at the disputed match but the gate-money received was nearly £15, and the entrance charge was a shilling. The visitors brought their own supporters with them. This was the largest gate for any Pentrediwaith football match I know of with the single exception of a charity match some years ago when Billy Meredith ceremonially kicked off. The gate on that occasion was nearly £30. The gates received at the other matches during Coronation Week were also good, if not as superlative as for the final, and out of a total income of £48 for the week there was a clear profit of £19. The question of the disposal of this sum soon shattered any hopes of future unanimity of village support for the football club.

Before this issue came up for discussion the annual general meeting already referred to was held and the new committee elected. What to do with the money was almost the first decision we had to make when we first met as a committee; it was preceded only by a revival of the discussion whether the team was to aim at heading the League with outsiders or struggle along with local talent. This discussion had a greater significance than I realised at the time. It was as the secretary, Charley (12), later pointed out, a personal attack on his own position.

The committee then began to discuss the disposal of the £19 profit made by its predecessors in the Coronation competition. I (4) had been told previously, as had the bank clerk chairman (1), that the disposal of these funds was not within the football committee's powers, the money having been promised to the Parish Council Coronation Fund. But when we both mentioned this the secretary (12) denied it was so. In view of the secretary's denial the committee decided to divide the money: three guineas were to be given to the Coronation fund, three guineas to the village hall where the meeting was being held and where the football committee always met, and three guineas to the British Legion Benevolent Fund; the rest was to be put at the disposal of three trustees to administer for the benefit of village charities.

There were heated arguments as to whether the Coronation committee and the Institute should have any money at all. These arguments were in each case first put by the English miner (7), although they did not represent his own personal views only but also reflected the animosity felt by many villagers towards the parish council, the Coronation committee and the Institute.

This meeting, like all Pentrediwaith committee meetings, was supposed to be held in secret. Nevertheless, the next day these decisions were known and discussed throughout the village. The chairman of the parish council was particularly angry at the turn of events. He actually produced Charley's (12) letter promising the money to the Coronation committee, and he wrote to the Football Association protesting. It was also known in the village who had taken which side in the committee's discussions and the English miner (7) was very unpopular for trying to prevent any of the money going to the Coronation committee. The code of reciprocity was invoked: his wife and child had attended the village Coronation Tea, and yet he wanted to spite the committee that had organised it.

Since the secretaryship of the supporters' club and of the Coronation committee (and later of the Carnival committee) was held by the same woman, the argument was put during the football committee's discussion that the money must go to the Coronation committee lest all the women be offended and the financial aid of the supporters be lost. This argument was repeated

in the village and helped to arouse the antagonism of the women to the football committee. The idea of being thus lightly bought and sold did not appeal to them.

Charley (12) countered these attacks by saying that the parish council had not fulfilled their part of the bargain since no one from that body or from the Coronation committee had come to help on the gate as had been promised. This was denied on the grounds that the manager of the Co-operative (2) had been on the gate twice and, although he was chairman of the supporters and a football club member, he was also on the Coronation committee. A more general point, and one which met with more sympathy in the village, was that even if the football club was in the wrong it was not right for the parish council chairman to take the quarrel outside the village by writing to the F.A. Indeed, for a very brief period this incident succeeded in uniting the football committee in opposition to the parish council chairman and his supporters.

While this controversy was at its height in the village, printed notices appeared in all the shops advertising a sacred concert to be given by a prize (brass) band on the Sunday preceding the Royal National Eisteddfod. This is a biennial event held in Pentrediwaith whenever the Royal National Eisteddfod is held in North Wales. In 1953, however, there was a departure from precedent and the posters were marked: 'In aid of the Football Players' Equipment Fund'. As this fund had never before been heard of, this started more talk in the village, not least by members of the football committee who shared in the general ignorance of what this was all about.

It was in this atmosphere of indignation about the Coronation money and doubt about the sacred concert that the new football committee met for the second time. The first item on its agenda was the appointment of trustees to administer the residue of the Coronation Competition funds. (This, of course, arose out of the decision of the prévious meeting to distribute the fund to certain named charities and devote the residue to general village charity.) The English miner raised the general question again in the light of his discovery that the parish council chairman had in his possession a letter from the football secretary (12) promising the profits to the Coronation fund. Members blamed the previous committee for not having handed the money over as soon as they had settled their expense accounts. It was finally decided to reverse the previous decision. All the money was now to be given to the treasurer of the Coronation committee with apologies for the delay and misunderstanding.

Some days after this meeting the chairman of the parish council and Charley (12), secretary of the football club, made up their differences. They agreed that the fault was with neither of them but was due to the slackness of the village bank manager who had been so slow in compiling the Coronation fund accounts. The final report on the Coronation committee's work sent to the local paper a fortnight later included the following paragraph in which the unity of Pentrediwaith was preserved in the eyes of the outside world:

> Special thanks are due to the local Football Club and their energetic Secretary in organising a knock-out medal competition which brought in over £19.

Thus ended the first major dispute in which the football club was involved in 1953–4. The second dispute started, however, even before the first had finished. At the same meeting of the football committee which reversed its previous decision on the Coronation fund, the question of the sacred concert was raised. The first move in this was made by Humphrey Biggs (6), who (like the principal woman agitator in the matter of the Coronation funds) had a long-standing quarrel with Charley (12). He said he had been told by a member of the Committee of the Welsh Football Association who, on a visit to the village, had happened to have tea in the café and hence had seen the poster, that the sacred concert would be against the rules of the Welsh F.A. In any case, Biggs asked: 'Who is running it?'

It appeared afterwards that the Football Association Committee member was fictitious. Like the minister at the parish meeting who threw the onus for his own views upon a former Member of Parliament, Biggs was using the F.A. member to shoulder some of the responsibility for Biggs's own views. I discovered afterwards that no one, with the exception of myself, had been deceived by Biggs's ruse, even at the time.

Charley (12) replied to Biggs's (6) question. He said that he was running the concert himself, with the aid of the football club treasurer, Timothy (3). Normally it was a British Legion affair but,

this being August Bank Holiday week, all the British Legion 'nobs' would be away. For this reason the British Legion had refused to sponsor it, and so he (Charley) was going to run it and give the money to the football club. He had not put 'Football Club' on the posters because it was a Sunday Sacred Concert,[7] and 'it didn't look nice to have a concert for the football on Sunday'. He could not have consulted the football committee as he had had to decide in a hurry, so he had formed his own committee consisting of himself and Timothy (3). The football committee decided that they would take no responsibility for the concert. After a warning against tittle-tattle from the chairman the meeting broke up.

When the committee next met, a week later, it was without Charley (12). He sent a letter to the chairman saying he felt bound to resign because of 'criticism inside and outside the committee'. In the intervening week he had had great difficulty in getting accommodation for the bandsmen, as people were refusing to 'help him line his pockets' and this, he felt, was the last straw. Humphrey (6) was elected to take his place despite his protests. I (4) was asked to be assistant secretary, but I refused to accept formal office although I agreed to continue to help with clerical work as I had done for Charley (12). The necessary contact between the old secretary and the new was also made through me as they were not on speaking terms.

Despite all difficulties the concert was duly held and was a success financially and socially. But it had a sequel. The evening before the football season was due to begin with a home match (in September 1953) the members of the football committee worked on the field preparing it for the next day. In the hotel Charley (12) gave a dinner for the players and committee of the previous season. He paid for this with the proceeds of the sacred concert. He committed two *faux pas* here. First, it was a festive occasion, although a boycott led to eight empty places and uneaten dinners out of the twenty-four ordered. Alcohol was in fact consumed, and songs were in fact sung, but it was not the debauch village gossip later made it out to be. Secondly, the dinner was not held in the inn owned by the president of the football club but in another pub. Later both these things told heavily against the football club, although it was not responsible for either.

Of the 1953–4 committee the vice-chairman, Percy (2), presided at the dinner. He did not partake of the drinks as he is a teetotaller. The treasurer (3) and one of the miners (9), who had both played the previous season, were also there. Two of the 1952–3 committee and myself (4) who happened to have gone to the pub for a drink also attended. There were five of the players present, and three outsiders staying in the hotel were invited. The other two present were Charley (12) himself and the chairman of the football club, Philip (1). The latter was a special case as he lived in the hotel permanently. When controversy later arose about attendance at the dinner, he claimed that he was merely having his usual meal a little late. He was, he said, just eating in the room, not attending the dinner. This did not deter him from playing the piano to accompany the songs.

The team which turned out next day included five of the diners and was defeated by nine goals to one. It had three more local players (eight) than had originally been chosen and two other changes. The season had got off to a bad start.

The following week the committee failed to choose a team at all on Monday for want of players and had to meet again later in the week. One group in the committee, two miner friends of Charley (9 and 10), put the blame for the bad start on Humphrey (6). Villagers who had been among the most active in seeking Charley's (12) removal were now agitating for his return. The next week another of the miner group, Edward (25), a friend of Charley's and brother of Edmond (9), agreed to play for the team and then failed to turn up at the bus. He was afterwards found to have played for another village. The first and only person to speak against him in the committee on account of this was his brother Edmond (9) (see Fig. 28.1).

At this stage Adam (10), leader of the miner group, signed two more players, and Humphrey (6) signed six. Tension was increased between them when one of Adam's new players was not picked to play. The other did play, but it was generally agreed, and even by Adam, that he was no good.

After this weekly disagreements between Adam (10) and the secretary (6) continued until the final explosion a month later. The conflict came to a head when Adam under the agenda item 'Match Report' said abruptly, and as the opening to the discussion: 'I'd like to know who is supposed to choose the team anyway, the committee or the secretary?' The secretary explained that

one of the players had let him down at the last moment and he had had to find a replacement. Adam asked why he had asked Dai Evans, a player from a distance, to play. The committee, Adam asserted with justice, had chosen Bill Davies as first reserve. It was well known to the committee that the secretary had opposed the choice of Bill Davies, whom he disliked:

Secretary: 'Bill Davies was going to watch a commercial match with Lloyd on Lloyd's motor-bike.'

Adam: 'Nonsense, man! He was waiting at the hall with his boots.'

Secretary: 'Well, I spoke to Lloyd [another reserve] and he said he was going out with Bill. I thought it was Bill Davies he meant. Perhaps it was Bill Jones.'

Adam: 'Well, I'm through. I'm leaving this committee.'

Secretary: 'Good riddance to you! You've only been a nuisance.'

The English miner (7) calmed Adam (10) down and restrained him from attacking the secretary. The encounter between Humphrey (6) and Lloyd had taken place in front of my house and I gave evidence about the conversation. The chairman, Philip (1), went round the committee one by one, asking each in turn: 'Do you accept the secretary's version?'

Edmond (9), the miner friend of Charley (12), was the first to answer. He was hesitant but would not give the lie direct. He replied 'yes', and the other five said 'yes' in turn; only Adam held out, saying: 'No, I'm through!' Bill Davies was fetched from the billiard room near by and questioned. He confirmed part of the secretary's story. Once more Adam was asked if he accepted the secretary's story. 'No,' he said again. No amount of persuasion would make him stay and off he went. Another miner, Edgar (11), brother to William (8) and the team captain, Lance, never attended the committee again after that evening. Although he did not come to the committee to give his views, he was said, like Adam, to think the secretary was too domineering about the team. The committee was now down to its hard core, and it remained in this state for most of 1953.

When Adam had gone, the English miner (7) explained that he worked with Adam and Adam's resignation was, in his opinion, the result of long-standing grievances. The English miner also proposed a resolution that the secretary should have discretion to get whomever he could to play, once he had exhausted the list of players and reserves chosen by the committee. This same English miner also suggested in private that Edmond (9), another miner, would be the next to go. As Fig. 28.1 on p. 545 shows, this prophesy was fulfilled, over two months later. Eventually the chairman (1) also withdrew, and the English miner (7) himself left the district. By mid-December 1953 there was an effective committee of three, the secretary (6), the treasurer (3) and the Co-operative Stores assistant (5).

On a visit I paid to Pentre in late January 1954 I found that the committee had been reinforced by the father and one of the brothers of the secretary. An attempt to get the 1952–3 season's committee to come back had failed. The football club was run informally by the three committee members who remained. Gates continued to diminish and, although village interest still ran high, village support was negligible.

Disputes do not long remain unresolved within the football club. For since the common interest in keeping football going is not sufficient, the committee splits and members resign. But while disputes in committee are solved by the resignation, or a less dramatic falling-off in attendance, of dissident members, the argument continues in the village. Eventually, as had happened by the beginning of 1954, an opposition to the football club grows up. There was a large number of men who would have nothing to do with the football club, and some who went even further and worked against it. The club's posters were torn from trees and players left at short notice to play elsewhere. There was a move, to be described below, to cut off sources of financial aid from the club. A parallel opposition was built up in the women's supporters' club. On a visit to Pentre in September 1954 when the football club had ceased to operate, I learned that although the supporters still had £25 in the bank, they were refusing to let the football committee have it even to pay their outstanding debts.

It seems to me that this sort of development is characteristic in the history of Pentre institutions and, I suspect, of that of similar villages. Efforts are made to avoid conflict, but once a breach is made patent, it spreads through the village. The face-to-face nature of social contacts, and the multiplicity of ties which close residence in an 'isolated' unit brings about, makes this spread inevitable. After a period, which in the case of the football club

was three years from the original difference over policy, the village becomes so divided that the particular activity cannot continue. In this case the activity perished altogether and was replaced by a carnival. At the 1954 carnival which I attended there were significant absences, both from participation and from the organising committee, which indicated that this activity had also already collected its enemies. Even brief conversations with villagers showed that it was in process of collecting more. There were also, as we have seen, indications that the same fate had overtaken previous activities. Once this process of losing support begins – and it is of course inherent at the beginning of any activity which replaces an old one – it is cumulative. Those who remain loyal to the football, for example, provide a nucleus of resentment to the carnival. As the opposition grows within the village, the activity is less likely to succeed in its aims. Thus, as the football committee grew defeatist in outlook and dwindled in numbers, it was less and less likely to field a winning team. Pentrediwaith's name appeared weekly at the bottom of the League table. This increased village hostility to the club committee and made the committee less efficient.

The carnival, like the football, is also reported in the local newspaper and has prestige value as a symbol of the village to the outside world. When, if my predictions are correct, it ceases to be an efficient symbol, the progress towards its inevitable downfall will be hastened.

Although conflict bedevilled the internal affairs of the football club, efforts were made to avoid it, and in this attempt strangers played their own familiar role. The English miner (7) introduced the actual resolutions most likely to split the committee internally, or to isolate it from other villagers. When conflict did break out, it was he (7) or I (4) who made the conciliatory moves. Similarly, when Charley (12) and the chairman of the parish council settled their differences it was the bank manager, an 'outsider' by class and place of birth, who was finally given the blame. The chairmen of the football club in the three seasons I have considered were all Welsh and Nonconformist, but nevertheless a little removed in outlook from the committee they led. In the first two seasons the chairmen were employers, although they worked alongside their men. In season 1953–4, the chairman was a bank clerk who, although he had played for the club and was a Welsh-speaking Chapel-goer, came from elsewhere. During this season the chair was taken in his absence by the English-speaking Co-operative Stores manager or by myself.

Although the differences of opinion which arose were on issues which were real, the individuals who clashed, Charley and Humphrey, Humphrey and Adam, Charley and Tom, were in fact motivated by loyalties and dislikes forged outside the committee or the club. In the case of the football club, conflict triumphed over cohesion, for although the football team symbolised village unity against the outside world, the organisation of football was a sectional interest of the men. Among the men only a small group were active, although many were interested. There was insufficient external pressure to make them carry on despite their differences.

CARNIVAL AT PENTREDIWAITH

The history of the 1953 carnival in the village went on simultaneously with that of the football club just related but, in contrast to it, was not confined to the men. But although the major conflict was between the organised group of men and the organised group of women, men and women were united not only by other ties external to the carnival situation, but also by their joint acceptance of a limited objective. Their attitude can be summarised by saying that they were united in a desire to make the carnival a financial success, and divided on the question of what to do with the money they hoped to make. Both their unity and their division were reinforced, as are all social relations in Pentrediwaith, by ties of loyalty and dislike carried over from other and seemingly irrelevant situations. The conflicts were partially concealed, and their bitterness mitigated, by forcing strangers to the group into positions of leadership where they had to take the blame for unpopular decisions.

Pentrediwaith Carnival was conceived after the Football Supporters' Club dance which formed part of the Coronation festivities. Coronation 'Queens' were much discussed in the national and local press in the early months of 1953, and at this dance a 'Coronation Football Queen' was chosen and invested with a sash. The secretary of the supporters' club, Betty (14), and the then football club secretary, Charley (12), decided to organise

a carnival later in the year at which the chosen 'Queen' could be 'crowned'. (Numbers in brackets after names refer to Fig. 28.2 on p. 553.) They did not, of course, just think of this idea *in vacuo*. Apart from the stirring events at Westminster, there is an annual carnival and 'crowning' in the local market town. In the next village down the valley the annual flower show is accompanied by the crowning of the 'Rose Queen', who is taken by lorry in procession from her village to Pentre and back. Furthermore, the composition of elaborately decorated tableaux on lorries and the making of fancy dress are popular pastimes in this part of Britain. Some Pentre women are particularly expert and compete in contests for the prettiest or funniest fancy dress at many local shows, either in person or through their children.

After the dance the supporters' club devoted itself to preparations for the carnival. I have already mentioned that in 1953 this club was an all women's organisation headed by a male chairman, Percy (2). He is in a very real sense outside the group he leads. We have already seen how the football committee is limited in the class from which it draws its membership because of the physical work and duties involved. The supporters' club is united into an abiding interest-group because its members work together in another way. During the preparations for the 1953 carnival, for example, the women concerned met together several times a week as a sewing-group preparing dresses for the 'Queen' and her retinue, the chairman being naturally excluded. Moreover, other village women known to have an interest in this sort of work were drawn in, including some who were openly hostile to the football club. These sewing-group meetings gave the women an opportunity for discussion and argument that was not paralleled in any male institution. Consequently the women were sometimes able to reach unanimity in their attitude to certain matters before they had ever been discussed in open committee. Other conflicts were resolved during the discussions of this informal work group and never came into the open at all.

The supporters' club committee and the extra women who were interested in the sewing preparations invited the football club committee to join with them in organising the carnival. They also invited a few men to join in order to carry out specific tasks. One of these was a Lancashire man, Mr Higgins (26), who kept a pet shop in the village and with his wife ran a kennels nearby. He was invited to organise a dog show. Another was Rhys (21), who had once been a footballer for the village but had been dropped from the team. Charley (12) and Adam (10) remained on the carnival committee after they had broken with football. Edward (25) continued to attend even after he had stopped playing for Pentre and started to play regularly for an English village. Other members of the football committee disapproved of the carnival and would have nothing to do with its organisation.

I (4) typed the original invitations to join this committee at Charley's (12) request before he resigned the secretaryship of the football club. I still possess the original copy in his handwriting. It reads:

Pentrediwaith F.C.
Carnival Committee

Dear Sir,

You have been selected a member of the above Committee. Kindly attend at the Hall on ..... at 7.30 p.m.

Charles .....

Hon. Sec.

I was not present at the first meeting of the carnival committee, which took place just after Charley's (12) resignation from the football club. This meeting took some decisions which afterwards assumed great significance. One of these was to accept Charley's offer to draft the posters, the programme and letters of appeal for funds. Charley was also asked to notify co-opted members of the football club and others of the next meeting. He had the letters of appeal typed by an English retired lady living in the village, instead of by me who had previously done his typing for him. Although no one realised the significance of this at the time, there was no reference to the football club either in these letters or in the note which I and other co-opted members received asking us to attend the carnival committee's second meeting. This was unheaded and read:

Dear Sir,

You are requested to attend a Carnival Committee on Tue. next August 25th 1953 at 7.30 at the hall.

It was signed with Charley's (12) initials. Neither the posters nor the programme which Charley drafted made any reference to the football club.

The committee met and made simple decisions covering the next two weeks; and the women worked and planned dresses and decorations outside the committee. The first signs that there was going to be open conflict in the relations of the football club and the carnival committee did not appear in the latter's proceedings, but in village gossip and at the football committee. Village gossip about a week before the carnival was due to be held was still concerned with the fateful sacred concert and the players' supper at the hotel. There was much to discuss, for not only had the money from a sacred concert been spent on beer, but also there had not been, nor could there be, any balance sheet. The reasons for the lack of published accounts were, first, that the concert had not been sanctioned by the Football Association and, secondly, that the entertainment had taken all the profit and there was none left to spend on equipment. It will be remembered that the concert was held in aid of the 'Football Players' Equipment Fund'. Because of these facts feelings against the football club committee ran high, although it was well known that the existing body had specifically dissociated itself from the concert and the dinner.[8]

In this atmosphere the football committee (see Fig. 28.1) held its weekly meeting, and at it a miner friend (9) of Charley's (12) asked, after the main discussion was over, if the committee realised that the money from the carnival was not to go to the football club at all but to the carnival committee for next year. He said, when asked, that Charley had told him that the carnival committee had decided this. I (4) said that it had not, in my knowledge, and Percy (2) said that he was chairman of the carnival committee, which was really just the supporters' committee and, as far as he was concerned, the money was for the football club. In the same week I was told that the question of where the profit was to go was receiving daily discussion in the women's sewing-group which was then working every afternoon.

Once the question had come into the open, there followed in the next fortnight two meetings of the carnival committee with a highly successful carnival in between. A week after the last of these meetings the football supporters' club held their annual general meeting – which should in fact have been held before the football season, but had been delayed because of the carnival. These three meetings and the carnival showed very clearly the division of interests between men and women in Pentrediwaith. They also provided the clearest example of the role of strangers in this society.

On the afternoon of the first of these committees, the last formal meeting before the carnival itself, the sewing-group met as usual and discussed the question of not giving any profit they made on the carnival to the supporters' club for football. The supporters' treasurer, Jane (15), told me on her return from this session that she was determined to raise the matter at the committee meeting that evening. The committee met in the main room of the hall with twenty-one people present (see Fig. 28.2). At a table in front of the stage, but on the same level as the rest of the committee, sat the chairman, Percy (2), flanked on either side by the treasurer, Jane (15), and secretary, Betty (14), of the supporters' club who were, of course, also treasurer and secretary of the carnival committee. Betty (14), the secretary, was a woman from the oldest part of the village on the hill beside the church; she had also been secretary of the Coronation committee; she was Welsh-born but English-speaking. The treasurer, Jane (15), was Welsh-speaking, Chapel and the wife of a local builder, Tom, who was the founder-chairman of the football club. At this committee meeting, as the diagram shows, members did not sit at random; they were grouped as kin, as cliques and as voting blocs of men against women, women against men. The one man who voted with the women and two other dissident men are (21), (12) and (25) in the diagram and, as can be seen, they are on the fringe of the men's group. All the committee members were villagers in the territorial sense except myself (4) and Mr Higgins (26); (8) and (11) were born in Pentre but lived in the next village.

Although they did not all act in this situation as kin, many members were closely related: (2) and (17) were husband and wife; (27) was the son of (15) and (3) was courting (15)'s daughter; I (4) lived at (15)'s house; the largest group linked by kinship were (19), (20), (21) and (28), and they acted and voted as a group; (8) and (11) were also related to these but acted independently (their relationship is shown below the main diagram); (19), (20) and (21) lived together on the main street of the village, and it was in a room of theirs that the sewing-group met. In addition (28) was the next-door neighbour, collaborator in handicrafts and friend of (18); (23) was actually the daughter of (18), but they

Secretary (14) Betty O
Chairman (2) Percy Δ
Treasurer (15) Jane O

Δ (8) William

Δ (11) Edgar

O (16) Mrs Morgan

Δ (10) Adam

O (17) Mrs Chairman

O (18) Mrs Price — Δ (27) Ian

O (19) Mrs Morgan — Δ (3) Timothy

(20) Mair O — (21) Rhys Δ — (12) Charley Δ — (4) Myself Δ

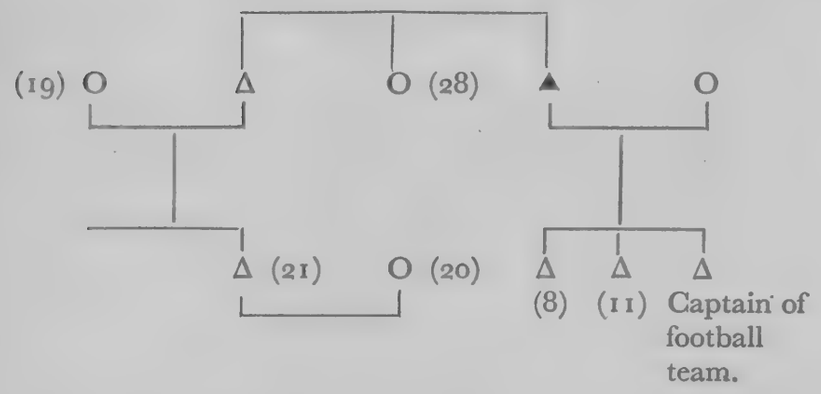
O (22) — Δ (26) Mr Higgins

O (23) — Δ (25) Edward

O (24)

Not present but active
O (28) Mrs Green

Fig. 28.2 First carnival committee



Carnival committee members' interrelationships

had quarrelled very seriously and were not on speaking terms; (22), (23) and (24) were close friends and attended football matches together; they did not share the sewing interests of the other women and were, in fact, very much younger than any of the others; (23) was also brother's daughter to (10). All the women except (22) were married; the three men who were not were either engaged or 'courting strong'; (12) and (25) were both miners in 1953, having been friends previously and having entered the mines when a bakery business in which they were partners failed; (10), (11) and (8) were members of what I have called the 'miners' group' of the football committee.[9]

When this committee met there was some general discussion; then Mrs Morgan (19) raised the question of the disposal of the surplus carnival funds. She said she had done a lot of work for it and was raffling a chair to raise money. She did not see why the football club should take all the money and the carnival have to start from scratch again the following year. She would not have helped to raise money just for the football club to waste. She had not realised it was in aid of the football. She formally proposed that a proportion of the profits be put aside for the carnival next year. When the chairman (2) called for discussion on this proposition, saying that he personally was opposed to it, Mrs Morgan (19) immediately withdrew her proposition, but the discussion continued. This happens frequently in Pentrediwaith: if a local proposer makes a suggestion he withdraws it as soon as it shows signs of raising controversy.

Most of the men present, including myself (4) and Charley (12), spoke against this plan, but Rhys Morgan (21) supported his mother (19). He said the football club would only waste the money. He had come on the committee to help at the carnival, not to help the football club. Nothing had been said on the programme or the posters about football. I was told afterwards that this branch of the Morgan family had quarrelled with the football club, two or three years previously, when after only one game Rhys (21) had been dropped from the team.

After a great deal more discussion, in which most took part and the same points were reiterated, Mr Higgins (26), an outsider to the village, intervened with great vigour. He pointed out that the football supporters' club had provided the initial finance and would stand the loss if there was one. It would be positively illegal

for them not to get the profit. He proposed: 'That all the profits go to the Football Supporters' Club.' This motion was put, and carried after a show of hands, by 11 notes to 9. All the men except Rhys voted for this motion. The only women who did were the chairman's wife (17) and another (and unrelated) Mrs Morgan (16) who had not, I was told later, been attending the sewing-group. Both the secretary (14) and the treasurer (15) of the supporters voted against their organisation receiving the money. This matter, it appeared at the time, being settled, the committee proceeded with its business. This was actually another dispute, but this time among the women only. It was a question as to which village children should attend the 'Carnival Queen'. The argument was carried on by the women who were anxious for this honour for their own children or grandchildren.

The women of the sewing-group who presented so united a front at the committee did not accept defeat, and the argument continued over tea and sewing for the next two days. Outside the committee meetings the women were not so united; (22), (23) and (24) insisted that the decision was right, and so did the treasurer of the supporters (15) although she had voted against it. The other women said that (22), (23) and (24) had not done any of the sewing, so had no right to an opinion.

As for the men, the women said that none of them had done anything except Charley (12). This was admitted by the men, but they countered with the true observation that although in the nature of things most of the preparation had been done by women, on the day itself the stalls and much of the administration would be run by men. Mair (20), wife of Rhys (21), announced that, whatever happened, the proceeds of the raffle for a chair, which she was conducting, were not going to go to the football.

But all the dissidents were united in support of Mrs Green's (28) view that the whole business was the fault of that Mr Higgins (26), a 'stranger', who had interfered in Pentre affairs. 'All strangers,' she said, 'ought to be shot!' And on that note the carnival itself was held.

A detailed account of the carnival is not necessary here. It is enough to note that, despite false starts and minor catastrophes, the carnival was held and enjoyed. Profits from raffles and competitions and the original football dance, together with the profits of the carnival and the carnival dance, came to nearly £60. The gross receipts were over £80. The unexpected size of the profits intensified the rivalry of the claimants, but things remained formally quiet until ten days later, when the winding-up meeting of the carnival committee was held. Only fourteen of the original members attended this. A plan of the seating at the start of the meeting is shown on Fig. 28.3.

The meeting started with Percy (2) as usual in the chair. He had, however, to leave early because of another engagement. Other affairs had occupied the village since the carnival and Percy did not, he said later, think that further argument about the disposal of the profits would take place that evening. Before he left the tone of the discussion was set. It began in discord. Betty (14) said she had forgotten to bring the minutes of the last meeting. When one of the men offered to fetch them she amended this; she had forgotten to write them. The important decision of the previous meeting was thus not recorded in writing. The chairman (2) gave a résumé from memory of what had taken place at the meeting. He referred to the great success of the carnival and went on to single out for praise the secretary (14) and Charley (12). This brought immediate protest from Mrs Price (18), who instead praised Mrs Morgan (19) who had let

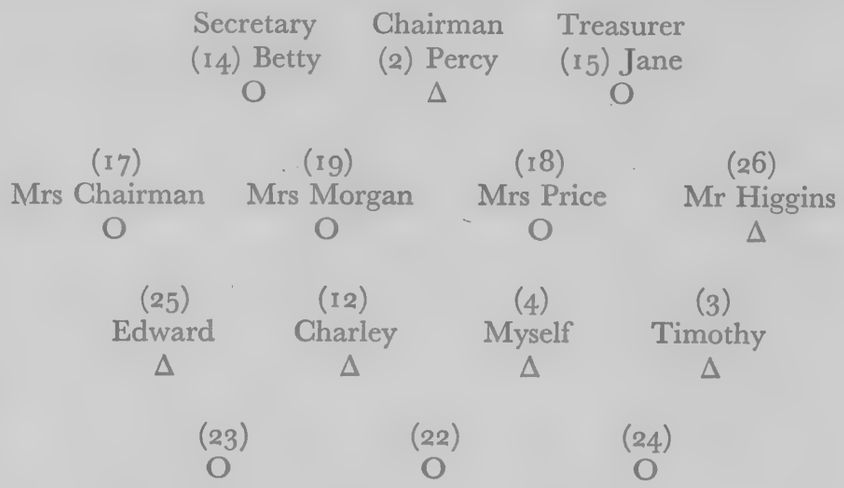


Fig. 28.3 Winding-up meeting of carnival

the sewing-group use her room, another villager who had lent a lorry and the chairman's enemy (not hitherto mentioned) who

had lent his house and his services for making paper-flowers.[10] The chairman apologised for his omissions and then tried to go, but he was 'persuaded' to stay and answer criticisms on his running of the children's sports at the carnival. They too were made by Mrs Price (18). This matter settled, Percy (2) announced that there was now only routine business of the supporters' club left and he went off to a rehearsal of his drama group. I myself was moved into the chair.

The question of who was to take the raffle tickets round the field the following Saturday was soon settled, and I asked for 'Any Other Business'. I did not expect any and was ready to close the meeting. But Charley (12) said there was just one point; he was not clear where the profit from the carnival was to go. To clarify things he wanted to put a motion to delete the previous committee's minute and substitute: 'That the money be divided fifty-fifty between the Football Supporters' and a Carnival Fund.'[11] He said that, after all, it had not been specified on the posters or appeals for funds that the carnival was connected with the football. Mr Higgins (26), the Lancashire man who had suffered much criticism for his previous stand on behalf of football and the letter of the law, put an amendment to Charley's motion: 'That £25 be given to the Football Supporters' and the rest kept by the Carnival.' No one would second either of these proposals and argument continued. The secretary, Betty (14), said that the supporters' had not financed the carnival at all. The treasurer (15) corrected her privately on this at the officers' table, occupied at the time only by myself (4) and these two. The correction was not audible to the main part of the committee, nor would the treasurer repeat it when I gave her an opening. She explained next day in conversation that the supporters had in fact financed the carnival to the extent of £11. She had not stated this openly in the committee because, as she said: 'I hope to live in the village all my life and I don't want to be at odds with my neighbours. I worked with the other women sewing and I know what they think about it.'

After more discussion in the committee, Mr Higgins (26) who, it will be remembered, had put an almost directly opposite motion ten days before, proposed that £30 be given to the carnival fund and that two trustees and a treasurer be appointed to administer this fund. His proposal was passed, by show of hands, unanimously and with acclamation. Betty (14), the secretary, and Mrs Price (18) were appointed trustees and Mrs Morgan (19) was made treasurer. I (4) adjourned the meeting amidst general affability and self-approbation. Mrs Morgan (19) and the supporters' treasurer (15) both said after the meeting that in future the carnival committee and the supporters would work together.

This general agreement did not long survive the adjournment of the committee, and members were involved in bitter arguments that evening. For example, the treasurer's (15) son, Ian (27), who had not attended this meeting, held that the decision was illegal; his mother (15) accepted this the next day, and said she would refuse to pay the cheque to the new carnival fund treasurer. During the next day every member of the committee that I met was claiming that she at least had not voted, and therefore could not be held responsible for the decision of others. Discussion continued in the village, but two days later, when Mrs Morgan (19), the new carnival treasurer, came to collect the cheque she was given it without argument. I gathered in the village that I (4) and, to some extent, Mr Higgins (26) were being blamed for the whole affair. It was asserted, probably with truth, that the proposal would not have been accepted if Percy (2) had stayed on to chair the meeting until its close.

Thus once more unpopularity incurred by making a decision which divided villagers was passed on to those it would least harm, and whose unpopularity had least effect on normal social relations within the village. In this case, Percy (2) who left early, either by accident or by design, avoided the criticism which, as a villager slightly different from the others, he usually received. His place was taken by Mr Higgins (26) and myself (4), two complete 'outsiders'. Despite this satisfactory apportionment of blame, it was expected that the controversy would be revived at the annual general meeting, now overdue, of the football supporters' club. This was open to the whole village. The sting had gone out of the argument, however, since the money had been paid over and no one expected that the trustees and the treasurer of the carnival fund would in any circumstances agree to its return.

The annual general meeting of the supporters illustrates well the division of opinion between men and women in Pentrediwaith. This had been clear throughout the controversy and had been exploited by Charley and the Morgans for their own ends. At the

general meeting the division was made completely explicit. Fig. 28.4 below gives the seating plan and attendance at this meeting. Percy (2) did not attend and his place was taken by Philip (1), the chairman of the football club, who had been away from the village on holiday during most of the carnival controversy and during the carnival itself.

Philip (1) opened the meeting by saying what a pleasure it was for him to preside over the supporters. This was especially so, because by their magnificent efforts in running a carnival, they had raised so much money for the football club. He was sorry to have been away and to have missed the carnival. He had heard that £30 of the sum raised was going to a carnival fund. Not, of course, that the football club begrudged the money; but, after all, the football supporters' had put up the original £7 (*sic*) and the football club would have suffered any loss incurred. Charley (12) 'begged to differ' about this and reiterated his now familiar point that the posters and the letters made no reference to the football club. Jane's (15) son, Ian (27), asked innocently who had drafted the notices. Jane, the supporters' treasurer (15), recalled that the carnival committee and the supporters had at the winding-up meeting of the carnival undertaken to work together in future. This recollection was not entirely accurate, as this undertaking was made only after the meeting had ended. But there were, of course, no minutes to verify or disprove it. Charley (12) and Timothy (3) denied it and Charley (12) appealed to me (4) to support him. On this occasion, however, after the week of sidelong glances which had followed my last intervention as chairman, I kept silent and would not comment one way or the other

Mrs Morgan (19), treasurer of the carnival fund, said they did not want to give all the money to the football club to spend, as the club might go out of existence and then they would have nothing to show for their carnival efforts. They were quite willing to hold the carnival next year in aid of the football again if the club still existed. Ian (27), who came to Pentre only on brief visits, said that that was all very well, but the football supporters had started a carnival, and now the situation was reversed. A carnival committee was very generously saying that they might give the football club a donation. It seemed sharp practice to him.

Charley (12) said that he would like to know why, while usually the supporters' treasurer (15) signed cheques on her own, the cheque paying £30 to the carnival fund had had to be signed by the chairman (2) as well. It appeared that what in fact had happened was that the bank had required a 'minute' opening the account, as is usual bank practice. This had built up in village rumour to the story given by Charley (12). This part of the discussion was closed when Philip (1) revealed that the football club was having difficulty in raising its usual donations, because

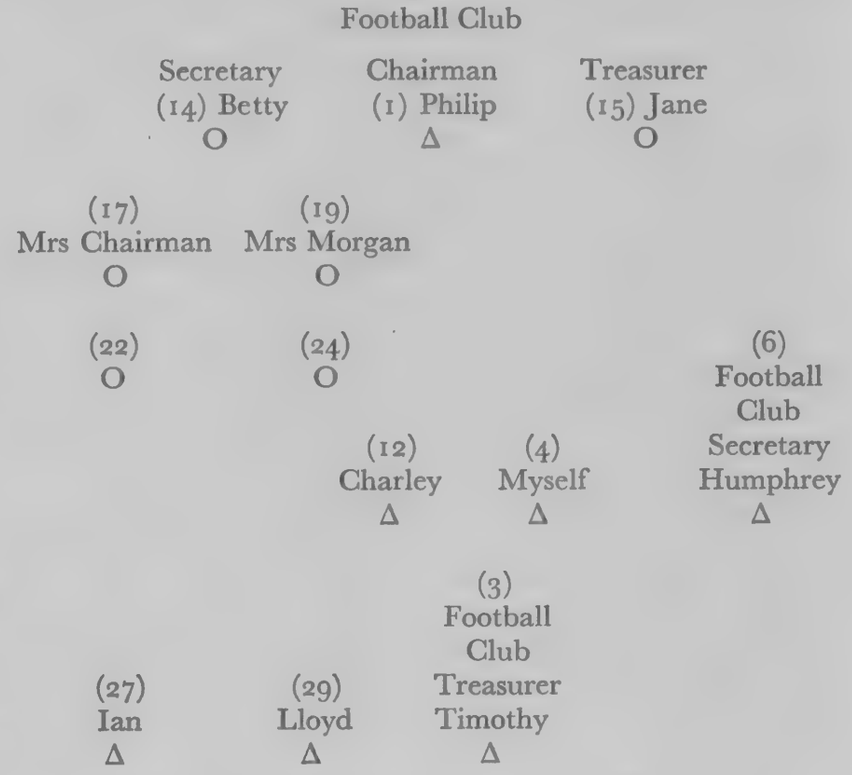


Fig. 28.4 Annual general meeting of supporters' club

the donors who had given to the carnival felt that they had done their bit, and would give no more to the club.[12] At this point in Philip's statement Charley (12), who was working the night-shift in the pit, had to leave. His departure was the signal for an almost audible relaxation of tension and Philip, amidst murmurs of

agreement, put squarely upon Charley's shoulders the blame for the admittedly bad relations between the football club and the majority of villagers. The incidents of the Coronation funds, the players' dinner and his alleged extravagance in bringing outside players to home matches by taxi, were all mentioned and unanimously condemned.

The 'footballers', however, were not to escape as lightly as that. Once the carnival quarrel had been disposed of by blaming it on Charley, the women had other things to speak of. The supporters' committee members took it in turn to detail the way in which the men failed to carry out, or even recognise, their reciprocal obligations. The cases they cited had in fact happened in previous seasons when other individuals had controlled the football committee. Jane (15), the supporters' treasurer, for example, told how she had asked a certain footballer three seasons previously, although she did not say so at the meeting, to buy a whist drive ticket. He had refused. She continued:

> 'What would you footballers do without the Supporters' Club?' I said.
>
> 'Huh, what would the Football Supporters' Club do with no one to support?' he said.
>
> 'Well,' I said, 'I always knew you were ignorant, but I didn't realise you were such a fool as that!'

Mrs Morgan (19) then told how, when she had been on the supporters', she had gone to an away match on the bus with the team. She had had to pay two shillings, whereas the football committee members had travelled free. This had also happened two years previously. (In 1953 football committee members paid their share with the rest.) The secretary (14) then raised the vexed question of the men's attendance at whist drives. Timothy (3), stung, retorted that someone had to pick a team, and if the supporters *would* arrange whist drives on Mondays when the football committee met, what did they expect? Betty replied: 'Fair play for Charley! He always comes, even on Monday.' Humphrey (6), the secretary of the football club, and Timothy (3), its treasurer, promised to be economical in future, and that all the football committee would try and come to whist drives.

This discussion was followed by a rapid election for the supporters' committee. The chairman, Percy (2), and the secretary, Betty (14), were re-elected. The treasurer, Jane (15), resigned on the grounds of ill-health. Aided by Ian (27), her son, she managed to escape the treasurership but not the committee. Mrs Morgan (19), already treasurer of the carnival, was persuaded to take the post for the supporters as well. This, largely the work of Philip (1), was an astute move,[13] as it not only brought in the group that she controlled, but it also resolved the feud between that group and the football club. In addition, it ensured that if the football survived, the carnival committee would co-operate with the supporters' club, since the officers of both were the same and fear of public opinion would prevent them favouring one club at the expense of the other; (17), (22) and (24) were re-elected in addition to (23) who had deliberately stayed away in the vain hopes of avoiding re-election. Finally, two young unmarried girls, one still at school, were added to the committee. It was explained that, since they always went to home matches because of their interest in the young men, they could be put to good use as raffle-ticket sellers. The elections completed, the meeting adjourned and so ended, at least for the moment, the controversy which surrounded the disposal of the carnival funds.

This account of the formal meetings concerning the carnival finances, and the relations of the carnival to the football supporters' club, gives the framework of the dispute. But the argument was not confined to the handful of villagers who attended these meetings. Nearly everyone in the village knew about it and discussed it. It was for a period almost the only topic of conversation in the pubs and in the shops. It divided households, at least all the families I knew best and visited regularly.

The protagonists of different points of view in the committees and at the meetings were spokesmen for the village. But these spokesmen were affected by many factors entirely unconnected either with the aims of the committees or the proceedings themselves. Decisions were taken and a compromise made after a complicated process of argument inside and outside the committees. The side that people took was determined not by an impartial objective judgement on the real issues involved, but by their own social position in the community as individuals and as members of particular groups. At the same time the division of interests between men and women dominated the discussions. Thus Charley (12) was motivated by his own relations to the

football club, as a result of its dispute with the parish council and also for more profound reasons already discussed. But he had to reconcile this attitude with his position as a man and as a member of the sectional group of miners. It will be remembered that he voted with the men at the first meeting. Jane, the supporters' treasurer (15), had to balance her own position as a woman among the other women of the sewing-group against her responsibilities as treasurer. She had the choice of annoying her fellow women, or the men of her own household and of the village.

The men in general were concerned with the continuance of football and ensuring that the club got the money it needed. The women's sewing-group wanted to ensure that they had means to continue with their hobby of making decorations and fancy dress. Their attitude was complicated by previous disagreements with the football club as individuals suffering personal slights, and as a group with interests opposed to those of the men.

At the crucial points in the public discussions, villagers like Mrs Morgan (19) and Charley (12) made suggestions which precipitated crisis, but they then withdrew into the background and allowed a stranger to come forward and put the formal motions which actually split the committee and village opinion. After the first two committee meetings a stranger was blamed, and although Charley (12) too was blamed in his absence at the third, villagers gave little public recognition to their own divisions. Their general verdict was that they could have managed all right, but that strangers *would* interfere and cause trouble.

Notes

1. It should not be thought that the handling of sheep-dogs is a skill practised only by shepherds. Pentrediwaith's great exponent of the skill was a quarryman. His son, who carries on the family tradition, works in a factory.

2. At the 1954 carnival, which I attended, it was announced that the carnival was in aid of a children's playing field, the football team having gone out of existence.

3. Alwyn Rees, *Life in a Welsh Countryside*, U.W.P., Cardiff, 1950, pp. 91 ff.; Huw Evans, *The Gorse Glen* (*'Cwm Eithen'*), Brython Press, Liverpool, 1948.

4. The costs of running a football club even on a small scale are surprisingly high. In 1952–3 the club had a turnover of over £260. In the previous season it had one of over £200. The figures for 1953–4 are not available.

5. According to J. Kempe, 'A Pilot Survey of Much Marcle', *The Sociological Review*, vol. 41 (1949), a similar problem arose in Much Marcle. This, however, was a flourishing community from the economic point of view and they solved it by having two teams. A similar solution was not possible in Pentrediwaith because of the economic situation.

6. One of the rules of the Welsh Football Association is that an amateur player cannot be a registered player for more than one club under its jurisdiction. There is therefore a struggle at the beginning of the season to persuade good players to sign for one's club before another club succeeds in 'signing' them. During the season it is possible to get players transferred from one club to another within the League, but only on a player's own initiative. It is also possible to sign players from clubs affiliated to other Football Associations. A club in the Shropshire F.A. succeeded in enticing the Pentre captain to play for them on a few occasions in 1953. Another Pentre player refused to play for Pentre while Biggs was in power and went to a Shropshire club of his own accord.

7. When the concert came off the 'sacredness' was not very apparent. There were fourteen items which included only two hymn tunes. Two others were Eisteddfod test-pieces and the remaining ten ordinary light brass band music.

8. In the same way at a general meeting of the supporters' club, the football committee was attacked for opinions expressed by a former committee member three seasons previously.

9. There were more distant relationships involved in this committee, though they were never mentioned in the discussion of events. I learned of them only through specific inquiries and in general discussion on the theme: 'All Pentre villagers are related.'

10. This man had once been a member of Percy's (2) 'Church' drama group. He had quarrelled with Percy and formed his own 'Chapel' drama group.

11. Although the official name is the Football Supporters' Club, it is always referred to simply as 'The Supporters'. There were some women who argued that, this being so, they could support anything they wanted, but this argument was not taken very seriously even by the anti-football faction within the club.

12. The football club's income by donation from its president, vice-presidents and the brewery companies is between £25 and £30. The carnival raised only £15 by direct donations. In fact, although Philip (1) did not mention it at this stage, the players' dinner by its snub to the publican–president had cost the club his substantial donation.

13. Or would have been if the football club had survived. Since writing this I have learned that it did not.

This excerpt by the British sociologist Percy Cohen is taken from his book 'Modern Social Theory', first published in 1968. Basically, Cohen says that there is no single satisfactory theory of social order. *He examines four types of theory which have been put forward by sociological theorists, and finally he concludes that 'All social order rests on a combination of coercion, interest and values.'*

THE PROBLEM OF ORDER

Percy Cohen

To assert that sociological theory centres around the problem of social order is to invite the charge of a conservative ideological bias; for to emphasise order is, on some views, to affirm its desirability. Although people in most societies, do, I think, desire order – they may not accept any kind of order at all costs – this is not a sufficient reason for placing the problem at the centre of theoretical inquiry. This is done, rather, for the following reasons: first, order is itself something positive and its opposites are only conceivable in terms of it; second, the very idea of human society presupposes order; third, the existence of social order is problematic and cannot be taken for granted; and fourth, the investigation of the problem of order illuminates the nature of disorder in its various aspects.[1]

ASPECTS OF SOCIAL ORDER

The term 'social order' can mean a number of things. The first meaning refers to the existence of *restraint*, the inhibition of impulse or, more specifically, to the control of violence in social life. The second meaning refers to the existence of *reciprocity* or *mutuality* in social life: the conduct of each individual is not random or haphazard but reciprocates or complements the conduct of others. The third meaning refers to the element of *predictability* in social life: men can act socially only if they know what to expect of one another. If these expectations are to be fulfilled then they must be fairly consistent with one another; the fourth meaning which can be attached to the notion of social order is that of *consistency*. The fifth meaning is that of *persistence*; there can be no predictability and consistency in social life unless its forms endure.

All of these meanings are both logically and empirically related. To say that men do not always act towards one another on impulse, but restrain themselves, is to imply either that they inherit their constraints genetically, or that they learn to internalise them during a process of social education. As far as we know men do not have genetically inherited social instincts; but they do have culture and the ability not only to be influenced by it but to create it.[2] An important part of any cultural heritage are the sets of reciprocal rights and obligations which men exercise in their dealings with one another. If men are to restrain themselves in these dealings they must also have some means of knowing the demands of others; and if they are to deny themselves certain courses of action they must have some recompense, in the form of mutual assistance or, at least, mutual non-interference. All of this implies the existence of rules or norms. Rules are, of their very nature, general; they cannot be made to fit any and every situation; indeed, their function derives from their generality, and generality implies predictability. If what A does is dependent upon what B does, then A must have some expectation of what B will do; and if what A chooses to do, or is forced to do, is affected by what B does, then A must have some idea of the relevance of B's actions for his own; and if the expectations of B are similarly dependent upon, or influenced by, those of A, then there will be certain common or corresponding expectations guiding the conduct of both A and B in certain situations, or types of situation. The establishment of norms increases predictability in social life by specifying a fairly limited range of expectations for particular types of occasion. But if predictability is to be maintained, the different sets of expectations in the same or

linked circumstances must be kept reasonably consistent with one another; or, putting it another way, the different norms, or sets of norms, must be integrated as a system. This is not to say that the different social norms must 'reflect' one another: the inequalitarian norms which govern or define the relations between parents and young children are in no way incompatible with the equalitarian norms which define the relationships between certain categories of adults in certain situations, even when these adults are also parents and children to one another. However, norms which accord to parents total control over the conduct of their children may well be incompatible with those which accord to the State certain rights over its subjects. If such conflicting norms exist then, in certain circumstances, men do not know what to expect; predictability and, indeed, reciprocity may break down. Such circumstances are common enough; but if they were typical of all social life there would be chaos.

Persistence is implied by predictability. This only means that on two or more successive occasions the same norms will prescribe conduct for the parties concerned. But persistence goes a good deal farther than that; in most societies *some* norms persist unchanged, or relatively unchanged, for several generations; and in some societies *most* of the norms persist for several generations. There are a number of reasons for this. First, some of the norms of family life, specifically those which govern the relations between the old and the young, actually exist to bridge the gap between the generations; although these may change with changes in the nature of this gap, they do not change entirely; for the older generation have learned how to deal with the younger only by internalising such norms from their own experience. Second, those who assume positions in existing social institutions do not create these institutions in their own image; there are no complete revolutions.

Each meaning of the term 'social order' is, in fact, an aspect of it; and each such aspect has its opposite. If men commonly suppress those drives, whims or impulses which go against the normative constraints, they also frequently allow them expression. While there is a good deal of reciprocity and even co-operation in social life, there is, equally, a good deal of opposition and conflict; indeed, reciprocity, co-operation, opposition and conflict may be compounded in the same relationships, as anthropologists and psychoanalysts have shown in their studies of the human family. While there is much that is predictable and consistent in social life, there are always some areas of uncertainty and inconsistency resulting from conflicts of principle, or conflicts between the expected and the possible. Finally, though all societies are in a process of constant change, this is more marked and more rapid in some circumstances than in others; and it always affects a given order of society unevenly; both persistence and change are characteristic of social life.

That societies simultaneously possess these opposed characteristics may be perfectly obvious. Yet it does not appear to be obvious to those who constantly proclaim that it is in the nature of societies to be riven by conflicts of interest and principle and to be in a constant process of change, while overlooking the fact that it is equally in its nature to be orderly and persistent.

EXPLANATIONS OF SOCIAL ORDER: THE COERCION THEORY

In general there have been four main types of theory to explain the existence of social order: the first has emphasised the factor of coercion, the second that of interests, the third that of values and the fourth that of inertia.

The coercion theory itself may emphasise the use, or threatened use, of physical coercion or the use of symbolic and moral coercion. According to this theory, order exists in society only, or largely, as a result of the power which some men have to command compliance from others; men do what is expected of them because they are compelled to do so by those who monopolise some means of coercion.[3] If they do not comply, they are threatened with some form of physical punishment, deprivation of property, resources or rights, or with some social stigma or supernatural sanction. This type of theory is often associated with a conspiracy theory which asserts that what occurs in society is the direct outcome of the wishes of some men who impress their will on others. But it is not necessarily wedded to such an assumption; for it may assert that those who command compliance do so in the name of all men.[4]

This theory can explain the various aspects of social order.

Men restrain their impulses and abide by norms because they fear the consequences of doing otherwise, or because they are constantly reminded of their moral need to do so by those who command them. They carry out their obligations and can expect others to do the same because a failure to do so will be punished by authority. Predictability is ensured in the same way. There must be consistency between different norms because if there were not this would come to the notice of those who command. And there must be persistence because it is, on the whole, in the interests of authority to maintain society in unchanging form for as long as possible; this makes their task easier and ensures continued power, prestige and other privileges for themselves.

The theory can also explain disorder, conflict and change. In all societies there are at least two sorts of conflict. First, there is conflict between men contending for positions of power: since order is based on nothing but coercion, any show of weakness on the part of the powerful will produce a struggle for succession among those close to power; this is most likely to occur when authorities are weak or under pressure. Second, there is conflict between the powerful and the powerless: whenever coercion fails or exceeds certain limits it produces a reaction against it to abolish it; these reactions are often successful when accompanied by struggles within the ruling group. During such periods of conflict all types of disorder occur: if there is no one able to uphold the norms they are not adhered to; the result is that 'instinct' and impulse re-emerge and chaos and violence ensue. In such conditions, social change may occur; though change may also occur if the powerful foresee such consequences: planned changes may result in greater repression or in less.

The coercion theory explains order in its various aspects and also explains the breakdown of social order and the occurrence of change. This is a great deal; but, for all that, the theory has serious weaknesses.

If it states that coercion is a *necessary* condition for social order, it can seemingly be refuted by examining those societies in which order exists without a single centralised authority which can coerce all of its members; these are the tribal societies which have been termed 'stateless' or 'acephalous'.[5] In such societies there are positions of authority within the domestic group or family, within the local community, within lineal segments or whole lineages; but there are scarcely any for the society as a whole; and commonly there are none even for the larger political units within the society. In these societies, the maintenance of order in the relationships between members of different segments is dependent upon a number of conditions. First, there is the potential use of force to settle disputes between members of different lineages or clans (or even larger units such as tribes) who can attempt to mobilise support from other members of the same social segment. Second, there are ties of various kinds, particularly those of marriage and extra-segmental kinship – if men marry outside of their lineages and clans this creates affinal ties for themselves and cognatic ties for their descendants – which cut across other kinship ties; these cross-cutting ties of kinship, or of ritual or political association, serve to restrain those who seek to mobilise force in the settlement of disputes; they also create a cohesive network of ties which, if strong enough, prevent the secession of segmental units from the larger whole. Third, there are sometimes intermediaries, who may have ritual status, who mediate between conflicting parties. An important aspect of the pattern of feuding is that internal segmentation within segments and the cross-cutting ties between them, tend to constrain the forces of organised violence; for men are reluctant to use force to seek redress of grievances unless they can mobilise support, since their opponents might meet their attack with strong defensive alliances; furthermore, nominal allies are reluctant to be drawn into states of feud with others whom they may wish to call upon as allies in some other disputes. In addition, various supernatural and moral sanctions set limits to the use of violence; and these are strengthened by rules which specify the social boundaries within which certain forms of redress are to be sought.

These ethnographic facts would appear to refute the argument that coercion is a necessary condition for social order. However, some counter-arguments should be considered. First, it is arguable that order in these societies *is* dependent upon coercive authority *within* each segment; for without this, segments would fragment into smaller units and the control of violence between segments would, in any case, not be possible without leadership. Second, it could be said that the potential use of coercion *between* units indicates that some form of it is necessary for social order, even if it is not exercised from a position of power. Third, it is clear that one of the constraints on social conflict is moral or symbolic

coercion; that is, the manipulation of ideas and moral attachments as a means of ensuring conformity or compliance.

The first of these arguments is partly convincing. A society cannot exist without some form of hierarchically organised coercion, even if this is based only on age or kinship seniority; however, this need not extend beyond the range of a segment of the society; and, for this reason, the theory, in its most general form, which postulates a necessary coercive force for the whole society, is rejected. The second argument also has something to commend it: the existence of countervailing coercion as a necessary condition of social order does suggest that some other form of coercion is necessary in the absence of a centralised form. (It could, of course, be said that this pattern of coercion is unstable by comparison with a centralised form: but there is little evidence to support such a view.) The third argument is largely unacceptable. Symbolic or moral coercion can only be exercised if men accept certain common values, and is, in the final instance, an inner coercion or constraint; and the effectiveness of inner constraints does not really support the theory that centralised coercion is a necessary condition for social order. However, it could well be argued, for two reasons, that there must be some form of hierarchically organised coercion resting on the potential use of physical force so that any form of moral or symbolic coercion can be effective: first, there must be some agent to apply sanctions when moral or symbolic influence is neglected, ignored or opposed; second, there must be some coercive agents, such as elders, who supervise the learning of moral standards and the internalisation of symbolic meanings.

One can agree that coercion is necessary without agreeing that it is also sufficient for the existence of social order. In the short run, the exercise of naked power may achieve some degree of social order; but in the long run it is bound to provoke opposition or the use of violence to overthrow it. And it is doubtful whether coercion as such is ever sufficient to maintain order even in the short run: the establishment of primitive and even of more complex states through coercion is commonly accompanied by dynastic struggles and sectional hostilities of a particularly violent kind.

If coercion is necessary, though not sufficient, for the maintenance of social order and, particularly, for the restriction of organised violence, then it would seem to be true that the weakening of coercive power is a sufficient condition for the occurrence of social disorder or change. This is difficult to test empirically for there is no society in which coercion disappears entirely. For example, after the removal of the Belgian presence in the Congo there was not a total absence of centralised coercion; there was a replacement of one coercive power by another; this encouraged rival claims to power, which was one of the most important factors leading to a breakdown of social order. But even though a rigorous test of the theory is not possible – at least with present techniques – there are still ways of assessing its explanatory power; for example, one could inquire whether the possible weakening of parental authority resulting from the development of peer-group associations, educational and other agencies and the legal protection of minors, did or did not contribute to family instability; or, to take a very different sort of example, one could inquire whether the impairment of the coercive power of feudal magnates did or did not facilitate the creation of a state bureaucracy. It seems reasonable to conclude that most evidence supports the view that the weakening of any form of coercive power contributes significantly to social disorder and to social change.

THE 'INTEREST' THEORY OF SOCIAL ORDER

The interest theory has two main variants. The first explains social order as resulting from a contract between men who find it in their interests to have some social arrangements.[6] This states that men cannot achieve their aims without the co-operation or, at least, the dependability of others. Since this dependability must be relatively predictable it requires the establishment of rules which accord to each party rights and obligations which are mutually beneficial, and which prevent some parties from benefiting excessively at the expense of others.

This theory can also account for disorder and change in the following way: when circumstances arise which are not covered by the existing rules then disorder and conflict follow, until, in time, the need for new rules is recognised.

There are two important defects to this type of theory. First, it explains social order largely in terms of men's intentions; and it is very clear that many social institutions exist without having been intended by anyone. Second, it fails to explain how these intentions can arise or affect social order prior to the existence of

some degree of order which would enable the parties to the contract to devise and establish it. These two objections are related: for it is difficult to see how men can have ideas expressing social intentions before they have the rudiments of social life, particularly language.

The second version of the 'interest' theory is far more subtle and complex than the first, and far more plausible. It states that social order results from the *unintended* consequence of the actions of many men separately pursuing their own interests; it is not that men discover that order is in their collective interests and then establish it, but that they establish it unwittingly and then, subsequently, discover that it is in their interests. The argument, stated more fully, is this: each man pursues his own interests and conducts himself in a way best designed to do this; in so doing, he finds that this inevitably brings him into contact with others, that he must adjust to the conduct of others, and therefore take it into account. This may lead, at first, to clashes. But gradually, each individual will find that his interests are better served, in the long run, by avoiding such clashes with some men, and possibly, by co-operating with them; this leads to the establishment of certain mutual expectations of conduct and the gradual modification of these until some balance is struck.

This theory, despite its defects, is one of the great contributions to the development of sociological theory; its merit lies just in this, that it does conceive of social phenomena in terms of causal processes which are, to some extent, independent of human will. It can explain the occurrence of social disorder and change with considerable economy: for the adjustment of interests is never *fully* achieved; there is always the possibility that individuals will develop new interests which will lead to conflicts with others and to pressures for new adjustments; and there is always the likelihood that circumstances will arise which could not have been foreseen and provided for in the establishment of any particular set of rules; and, finally, man's 'nature' is such that he will not rest content with the pursuit of particular goals, and his interaction with others may even stimulate or compel him to engage in new activities. Man is, on the whole, capable of rational action in pursuing his goals – though he is also considerably affected by sentiments and passions – but he does not rationally design the form of his social life.[6]

There are two main defects to this theory. The first, which has been exposed by Émile Durkheim[7] and later, following him, Talcott Parsons,[8] is that it does not explain the derivation of interests. Durkheim and Parsons both argue that these are, to some extent, given by society itself; that is, their existence presupposes social order. A defendant of this theory could argue that some interests are given by man's 'nature', by his condition in the natural world and by the bare fact of his coexistence with others; therefore his primary interests depend not upon society but upon nature. This argument could be used to account for some orderliness in the social behaviour of non-human primates, but it cannot account for the type of order which exists in human society, for this at least presupposes linguistic communication; and the rules of language presuppose society.

The second objection to the theory is that it assumes that men adjust to the conduct of one another with almost perfect freedom, their choices being limited only by nature and by the need to take note of one another's conduct. No consideration is given to the power which some men have to determine or influence the manner in which others will take account of their own wishes: in short, the theory so stresses the unintended consequences of actions that it overlooks the degree to which *some* intentions are more significant than others, even when the outcome is not what was intended by any one or more parties.

The process of adjustment of interests is never sufficient for the establishment and maintenance of social order; but it is necessary. Of course, to say this is almost tautological: it is always possible to show that a given order of society expresses some mutual adjustment of interests. Nevertheless, the theory has considerable value in directing attention to an important aspect of social order.

THE VALUE-CONSENSUS THEORY OF SOCIAL ORDER

The third theory states that order is based on some minimal consensus on certain values, which are predominantly moral but may also be technical and aesthetic.[9] The argument underlying this theory is that if men are committed to the same values they recognise a common identity as against others; they accept common goals and also certain prescriptions and prohibitions concerning the means whereby these goals can be achieved. The choice of

conduct made by each individual will be largely governed by these common standards and by the psychological satisfaction which is obtained from such conformity. The commitment to values enables men to devise means for reconciling or adjusting conflicting interests, and for turning coercive force into legitimate authority.

This theory can explain disorder and change in the following way. First, no man is totally committed to common standards and some, because of their upbringing, are never fully committed. When circumstances change radically, many men will abandon their commitments; but when they change even slightly some men will do this more readily. Thus, there are always individuals who are ready to deviate, and they are large in number when circumstances encourage this. The second possible cause of disorder and change occurs when there is a clash of values due to contact between different societies or when new values emerge which are incompatible with the old.

This theory which has, of late, been well favoured is hardly new; it was advanced by Auguste Comte and, to some extent, by Emile Durkheim who also recognised its weaknesses as a general theory. Comte argued that disorder in society was due to a lack of consensus on certain fundamental ideas and principles concerning the desired type of society and the proper means of administering it. He considered that consensus was reduced by the growth of the division of labour, which promoted sectional differences and conflict.[10]

Durkheim rejected much of Comte's argument. He considered that a unified system of ideas and morals was possible only in a relatively simple, undifferentiated society. He recognised that the division of labour promoted differences and conflicts, but considered that it also provided a new moral conception of interdependence. In highly differentiated societies there could be no consensus based on detailed moral rules and the acceptance of total systems of ideas; there could, however, be a consensus on certain diffuse moral values which prescribed limits within which different sets of rules could exist.

If this theory states that consensus is a *sufficient* condition for the creation and maintenance of social order then it is rather weak. In the first place, it fails to explain how some consensus can be reached without social order. Secondly, it is also false as an explanation of the continuity of order: for there may be widespread consensus in modern society on the desirability of higher living standards; but this is as likely to provoke conflict as to solve it. If there were not this consensus – if different social sectors had different values and only *some* desired higher living standards – then there might be far greater social stability.

But if the theory states that some degree of consensus or commitment to common values is a *necessary* condition for social order, it is far more acceptable. In a trivial sense the theory is almost indisputable; for human social life is simply inconceivable unless men have some common standards and are committed to maintaining them. In a more significant sense, the theory states that, underlying the acceptance of some common rules is the commitment to certain broad principles concerning the desirability of these rules. This implies – indeed, the very term 'value' suggests this – that there is a possibility of choosing between one set of principles and another. This is questionable: for while it may be true in some complex civilisations that men debate different and even opposed views concerning the desirable society and can to some extent choose between different courses of action in terms of these, it is not true at all of most simple societies; in these latter, in so far as there are choices, they are made within a common conception of social order which is relatively unquestioned. Where then does the commitment to common values lie? In those societies where there is a possibility of choice, where consensus is reached with difficulty and can therefore be said to reflect a commitment to one thing rather than to another? Or does it lie where no fundamental disagreement is possible? This is not mere sophistry. The question is aimed to elucidate the following point: if there is fundamental agreement on the nature of society and no alternative is conceivable, is it not trivial or truistic to say that this is based on commitment to common values? Is social order not simply equated with the commitment to common values?

There is a counter-argument to this. It states that the condition of value-commitment, which cannot be directly observed, and can only be inferred from certain objective indices, is phenomenally independent of social order itself. For example, it could be said that Hindu beliefs and values provide a set of common principles which define or justify the different rights and obligations of caste groups, and that this is a necessary, if not a sufficient, condition for main-

taining both an orderly relationship between occupational groups and a stable structure of power and prestige. What this argument implies is that without such an underlying commitment, which may be diffuse and even subliminal, men would not necessarily conform to the specific rules of social life but might well choose courses of action which would result in chaos. The difference between simple and complex societies would lie in this: that in the former the alternative would be simply between order and chaos, while in the latter it would also be between different *kinds* of order.

Despite its weaknesses, this theory is not easily dismissed: even if it does approach the tautological, it is of value in directing attention to an important factor in social life. And, as I hope to show later, it can be usefully combined with some elements in the other theories. Its unpopularity in some circles is due largely, and sometimes justifiably, to the inflated claims which have been made for it in one form or another. Comte and Parsons have supported it strongly as an explanatory theory; while men as different as Comte, Mannheim and Marx have held rather Utopian notions concerning the possibility of establishing a complex society on the basis of a vast consensus. Critics have been quick to point out that a commitment to common values and ideas in complex societies is unlikely to be extensive even where it is powerful.

THE 'INERTIA' THEORY OF SOCIAL ORDER

The fourth and last type of theory of social order is rather different from all the others in that it seeks only to explain one aspect of social order, namely the element of continuity or persistence in social life. It is also different in that it does not refer to a single factor or process but to any number of them, and does not necessarily compete with any of the other theories. The theory asserts, in brief, that if social order exists it provides the conditions for its own perpetuation. This sounds absurd: either it is tautological, in that the notion of social order implies that of continuity; or it is clearly false, for social order can break down or change, and to say that it sometimes persists is to revert to platitude.

Despite its seeming absurdity the theory has, when formulated more clearly and precisely, much to commend it; for it emphasises the point that some of the causal processes of social phenomena are often circular. What this theory *should* state is that when social order is maintained by a number of mutually reinforcing processes, it tends to resist pressures for disruption and change, at least from within.

This theory can be combined with any one of the others: coercion, interests and value consensus can each be introduced as factors within the theory or model which uses the assumption of inertia or equilibrium. In fact, all three elements can be combined within a theory or model of an ongoing social order.

CONCLUSION

A survey of these four types of theory brings one inescapably to two important conclusions. The first is that none of these theories can really explain the *origins* of social order; indeed, this may be an impossible task to carry out in purely sociological terms. (Whether an explanation in partly biological terms is possible, I do not venture to say.) The coercion theory is implausible because it presupposes that those who wield power invent a number of social norms. The contractual theory presupposes the existence of some form of social order within which a contract can be designed and agreed to; it therefore assumes, as part of the explanation, what is to be explained. The second interest theory comes nearest to success, for it assumes a long process of undesigned development from a state of disorder to one of order; but it cannot fully account for the existence of interests without showing that these depend on the existence of society. The consensus theory likewise presupposes the existence of a social order which can promote commitment to certain underlying common values. Finally, the inertia theory makes no claim to explain the *origin* of social order.

The second inescapable conclusion is that *each* of these theories contributes to the explanation of how social order *persists* – rather than how it originates – and how it breaks down and changes. For each theory states a *necessary*, though not a sufficient condition, for the continuity of any social order, once it exists. All social order rests on a combination of coercion, interest and values. This does not mean that every type of social order is dependent, to the same extent, on each of the three factors. In fact, they differ very much in the extent to which they emphasise these different elements.

Similarly, though all societies are characterised by some degree

of inertia, they differ in the degree to which this is the case. To understand why they differ in this respect – that is, to understand why some societies are more resistant to change than others – one has to investigate more closely the processes which are said to be interrelated in social systems.

Notes

1. Talcott Parsons, *The Structure of Social Action* (Free Press of Glencoe, Illinois, 1949).

2. Culture is, of course, not only negative, it does not merely constrain 'animal' behaviour, but transforms it and adds to it; it does not merely inhibit motivation but creates new motivations.

3. It should be noted that the English word 'order' can also mean 'command'; the confusions of etymology may well reflect popular social theories.

4. E.g. Thomas Hobbes, *Leviathan*, ed. Michael Oakeshott (Oxford, 1946) esp. pp. 107–8.

5. E.g. John Middleton and David Tate (eds), *Tribes without Rulers* (London, 1958).

6. Friedrich A. Hayek, *Individualism and Economic Order* (London, 1949) pp. 6–13.

7. Émile Durkheim, *The Rules of Sociological Method*, trans. Solovney and Mueller; ed. George E. G. Catlin (Free Press of Glencoe, Illinois, 1950) pp. 2–3.

8. Talcott Parsons, op. cit., p. 460.

9. K. Davis, *Human Society* (New York, 1959), pp. 143–4.

10. Auguste Comte, *Cours de Philosophie Positive*, vol. 4, pp. 429–31.

11. Émile Durkheim, *The Division of Labour in Society*, trans. George Simpson, (Free Press of Glencoe, Illinois, 1947) esp. pp. 364–71.

The following selection from Robert Merton is an extract from a long essay called 'Manifest and Latent Functions'. It was originally published in the 'American Journal of Sociology' in 1945. The 'corrupt' political machines to which Merton is referring are those which operated in many large American cities during the Prohibition era and the depression of the 1930s. In this period a much higher proportion of Americans than today were either themselves foreign-born or were children of people born abroad – many in eastern or southern Europe.

Although Robert Merton's essay has long been a 'classic' in sociology, it is included here with the intention of showing some of the weaknesses – as well as the strengths – of traditional 'functional analysis'. One of its strengths was to indicate reasons why such political machines persisted and how they were held together by the various 'functions' they performed. However, one may also question Merton's claim to have substituted sociological analysis for naïve moral judgements. Merton equally naïvely, one might argue, accepts the status quo*; he probably overestimates the benefits of the machines, even for such minority groups as Italian immigrants; he probably underestimates the suffering involved, as well as the part played by violence of various kinds. There is a lack of hard data of any kind, and a considerable element of ambiguity in the use of key terms such as 'functional' or 'latent'.*

Lewis Coser's piece – indicating that conflict may well produce positive benefits for society – has perhaps stood the test of time better (it was first published in 1956). It is also helpful in showing that stability and change are basically 'all part of the same thing' – namely a living society – and are by no means polar opposites.

Coser's analysis can be – and has been – applied to such situations as urban race riots. Racial conflict and violence perpetrated by extremists may nevertheless enable 'moderate' protest leaders to achieve some of their more limited goals.

MANIFEST AND LATENT FUNCTIONS

Robert Merton

SINCE moral evaluations in a society tend to be largely in terms of the manifest consequences of a practice or code, we should be prepared to find that analysis in terms of latent functions at times runs counter to prevailing moral evaluations. For it does not follow that the latent functions will operate in the same fashion as the manifest consequences which are ordinarily the basis of these judgements. Thus, in large sectors of the American population, the political machine or the 'political racket' are judged as unequivocally 'bad' and 'undesirable'. The grounds for such moral judgement vary somewhat, but they consist substantially in pointing out that political machines violate moral codes: political patronage violates the code of selecting personnel on the basis of impersonal qualifications rather than on grounds of party loyalty or contributions to the party war-chest; bossism violates the code that votes should be based on individual appraisal of the qualifications of candidates and of political issues, and not on abiding loyalty to a feudal leader; bribery and 'honest graft' obviously offend the proprieties of property; 'protection' for crime clearly violates the law and the mores; and so on.

In view of the manifold respects in which political machines, in varying degrees, run counter to the mores and at times to the law, it becomes pertinent to inquire how they manage to continue in operation. The familiar 'explanations' for the continuance of the political machine are not here in point. To be sure, it may well be that if 'respectable citizenry' would live up to their political obligations, if the electorate were to be alert and enlightened; if the number of elective officers were substantially reduced from the dozens, even hundreds, which the average voter is now expected to appraise in the course of town, county, state and national elections; if the electorate were activated by the 'wealthy and educated classes without whose participation', as the not-always democratically oriented Bryce put it, 'the best-framed government must speedily degenerate'; if these and a plethora of similar changes in political structure were introduced, perhaps the 'evils' of the political machine would indeed be exorcised.[1] But it should be noted that these changes are often not introduced, that political machines have had the phoenix-like quality of arising strong and unspoiled from their ashes, that, in short, this structure has

exhibited a notable vitality in many areas of American political life.

Proceeding from the functional view, therefore, that we should *ordinarily* (not invariably) expect persistent social patterns and social structures to perform positive functions *which are at the time not adequately fulfilled by other existing patterns and structures*, the thought occurs that perhaps this publicly maligned organisation is, *under present conditions*, satisfying basic latent functions.[2] A brief examination of current analyses of this type of structure may also serve to illustrate additional problems of functional analysis.

SOME FUNCTIONS OF THE POLITICAL MACHINE

Without presuming to enter into the variations of detail marking different political machines – a Tweed, Vare, Crump, Flinn, Hague are by no means identical types of bosses – we can briefly examine the functions more or less common to the political machine, as a generic type of social organisation. We neither attempt to itemise all the diverse functions of the political machine nor imply that all these functions are similarly fulfilled by each and every machine.

The key structural function of the Boss is to organise, centralise and maintain in good working condition 'the scattered fragments of power' which are at present dispersed through our political organisation. By this centralised organisation of political power, the boss and his apparatus can satisfy the needs of diverse subgroups in the larger community which are not adequately satisfied by legally devised and culturally approved social structures.

To understand the role of bossism and the machine, therefore, we must look at two types of sociological variables: (1) the *structural context* which makes it difficult, if not impossible, for morally approved structures to fulfil essential social functions, thus leaving the door open for political machines (or their structural equivalents) to fulfil these functions, and (2) the subgroups whose distinctive needs are left unsatisfied, except for the latent functions which the machine in fact fulfils.[3]

STRUCTURAL CONTEXT

The constitutional framework of American political organisation specifically precludes the legal possibility of highly centralised power and, it has been noted, thus 'discourages the growth of effective and responsible leadership'. The framers of the Constitution, as Woodrow Wilson observed, set up the check and balance system 'to keep government at a sort of mechanical equipoise by means of a standing amicable contest among its several organic parts'. They distrusted power as dangerous to liberty: and therefore they spread it thin and erected barriers against its concentration. This dispersion of power is found not only at the national level but in local areas as well. 'As a consequence,' Sait[4] goes on to observe, 'when *the people or particular groups* among them demanded positive action, no one had adequate authority to act. The machine provided an antidote.'[5]

The constitutional dispersion of power not only makes for difficulty of effective decision and action but when action does occur it is defined and hemmed in by legalistic considerations. In consequence, there developed 'a much *more human system* of partisan government, whose chief object soon became the circumvention of government by law. . . . The lawlessness of the extra-official democracy was merely the counterpoise of the legalism of the official democracy. The lawyer having been permitted to subordinate democracy to the Law, the Boss had to be called in to extricate the victim, which he did after a fashion and for a consideration.'[6]

Officially, political power is dispersed. Various well-known expedients were devised for this manifest objective. Not only was there the familiar separation of powers among the several branches of the government but, in some measure, tenure in each office was limited, rotation in office approved. And the scope of power inherent in each office was severely circumscribed. Yet, observes Sait in rigorously functional terms, 'Leadership is necessary; and *since* it does not develop readily within the constitutional framework, the Boss provides it in a crude and irresponsible form from the outside.'[4]

Put in more generalised terms, *the functional deficiences of the official structure generate an alternative (unofficial) structure to fulfil existing needs somewhat more effectively*. Whatever its specific historical origins, the political machine persists as an apparatus for satisfying otherwise unfulfilled needs of diverse groups in the population. By turning to a few of these subgroups and their characteristic needs, we shall be led at once to a range of latent functions of the political machine.

FUNCTIONS OF THE POLITICAL MACHINE FOR DIVERSE SUBGROUPS

It is well known that one source of strength of the political machine derives from its roots in the local community and the neighbourhood. The political machine does not regard the electorate as an amorphous, undifferentiated mass of voters. With a keen sociological intuition, the machine recognises that the voter is a person living in a specific neighbourhood, with specific personal problems and personal wants. Public issues are abstract and remote; private problems are extremely concrete and immediate. It is not through the generalised appeal to large public concerns that the machine operates, but through the direct, quasi-feudal relationships between local representatives of the machine and voters in their neighbourhood. Elections are won in the precinct.

The machine welds its link with ordinary men and women by elaborate networks of personal relations. Politics is transformed into personal ties. The precinct captain 'must be a friend to every man, assuming if he does not feel sympathy with the unfortunate, and utilising in his good works the resources which the boss puts at his disposal'. The precinct captain is for ever a friend in need. In our prevailingly impersonal society, the machine, through its local agents, fulfils the important social *function of humanising and personalising all manner of assistance* to those in need. Food-baskets and jobs, legal and extra-legal advice, setting to rights minor scrapes with the law, helping the bright poor boy to a political scholarship in a local college, looking after the bereaved – the whole range of crises when a feller needs a friend, and, above all, a friend who knows the score and who can do something about it – all these find the ever-helpful precinct captain available in the pinch.

To assess this function of the political machine adequately, it is important to note not only that aid *is* provided but *the manner in which it is provided.* After all, other agencies do exist for dispensing such assistance. Welfare agencies, settlement houses, legal aid clinics, medical aid in free hospitals, public relief departments, immigration authorities – these and a multitude of other organisations are available to provide the most varied types of assistance. But in contrast to the professional techniques of the welfare worker which may typically represent in the mind of the recipient the cold, bureaucratic dispensation of limited aid following upon detailed investigation of *legal* claims to aid of the 'client' are the unprofessional techniques of the precinct captain who asks no questions, exacts no compliance with legal rules of eligibility and does not 'snoop' into private affairs.[7]

For many, the loss of 'self-respect' is too high a price for legalised assistance. In contrast to the gulf between the settlement house workers who so often come from a different social class, educational background and ethnic group, the precinct worker is 'just one of us', who understands what it's all about. The condescending lady bountiful can hardly compete with the understanding friend in need. In *this struggle between alternative structures for fulfilling the nominally same function* of providing aid and support to those who need it, it is clearly the machine politician who is better integrated with the groups which he serves than the impersonal, professionalised, socially distant and legally constrained welfare worker. And since the politician can at times influence and manipulate the official organisations for the dispensation of assistance, whereas the welfare worker has practically no influence on the political machine, this only adds to his greater effectiveness. More colloquially and also, perhaps, more incisively, it was the Boston ward-leader, Martin Lomasny, who described this essential function to the curious Lincoln Steffens: 'I think,' said Lomasny, 'that there's got to be in every ward somebody that any bloke can come to – no matter what he's done – and get help. *Help, you understand; none of your law and justice, but help.*'[8]

The 'deprived classes', then, constitute one subgroup for whom the political machine satisfies wants not adequately satisfied in the same fashion by the legitimate social structure.

For a second subgroup, that of business (primarily 'big' business but also 'small'), the political boss serves the function of providing those political privileges which entail immediate economic gains. Business corporations, among which the public utilities (railroads, local transportation and electric light companies, communications corporations) are simply the most conspicuous in this regard, seek special political dispensations which will enable them to stabilise their situation and to near their objective of maximising profits. Interestingly enough, corporations often want to avoid a chaos of uncontrolled competition. They want the greater security of an economic czar who controls, regulates and organises competition,

providing that this czar is not a public official with his decisions subject to public scrutiny and public control. (The latter would be 'government control', and hence taboo.) The political boss fulfils these requirements admirably.

Examined for a moment apart from any moral considerations, the political apparatus operated by the Boss is effectively designed to perform these functions with a minimum of inefficiency. Holding the strings of diverse governmental divisions, bureaus and agencies in his competent hands, the Boss rationalises the relations between public and private business. He serves as the business community's ambassador in the otherwise alien (and sometimes unfriendly) realm of government. And, in strict businesslike terms, he is well paid for his economic services to his respectable business clients. In an article entitled 'An Apology to Graft', Lincoln Steffens suggested that 'Our economic system, which held up riches, power and acclaim as prizes to men bold enough and able enough to buy corruptly timber, mines, oil fields and franchises and "get away with it", was at fault.'[8] And, in a conference with a hundred or so of Los Angeles business leaders, he described a fact well known to all of them: the Boss and his machine were an *integral part* of the organisation of the economy. 'You cannot build or operate a railroad, or a street railway, gas, water, or power company, develop and operate a mine, or get forests and cut timber on a large scale, or run any privileged business, without corrupting or joining in the corruption of the government. You tell me privately that you must, and here I am telling you semi-publicly that you must. And that is so all over the country. And that means that we have an organisation of society in which, *for some reason*, you and your kind, the ablest, most intelligent, most imaginative, daring, and resourceful leaders of society, are and must be against society and its laws and its all-around growth.'[9]

Since the demand for the services of special privileges are built into the structure of the society, the Boss fulfils diverse functions for this second subgroup of business-seeking-privilege. These 'needs' of business, as presently constituted, are not adequately provided for by conventional and culturally approved social structures; consequently, the extra-legal but more-or-less efficient organisation of the political machine comes to provide these services. To adopt an *exclusively* moral attitude towards the 'corrupt political machine' is to lose sight of the very structural conditions which generate the 'evil' that is so bitterly attacked. To adopt a functional outlook is to provide not an apologia for the political machine but a more solid basis for modifying or eliminating the machine, *providing* specific structural arrangements are introduced either for eliminating these effective demands of the business community or, if that is the objective, of satisfying these demands through alternative means.

A third set of distinctive functions fulfilled by the political machine for a special subgroup is that of providing alternative channels of social mobility for those otherwise excluded from the more conventional avenues for personal 'advancement'. Both the sources of this special 'need' (for social mobility) and the respect in which the political machine comes to help satisfy this need can be understood by examining the structure of the larger culture and society. As is well known, the American culture lays enormous emphasis on money and power as a 'success' goal legitimate for all members of the society. By no means alone in our inventory of cultural goals, it still remains among the most heavily endowed with positive affect and value. However, certain subgroups and certain ecological areas are notable for the relative absence of opportunity for achieving these (monetary and power) types of success. They constitute, in short, sub-populations where 'the cultural emphasis upon pecuniary success has been absorbed, but where there is *little access to conventional and legitimate* means for attaining such success. The conventional occupational opportunities of persons in (such areas) are almost completely limited to manual labour. Given our cultural stigmatisation of manual labour,[10] and its correlate, the prestige of white-collar work, it is clear that the result is a tendency to achieve these culturally approved objectives *through whatever means are possible*. These people are on the one hand 'asked to orient their conduct towards the prospect of accumulating wealth [and power] and, on the other, they are largely denied effective opportunities to do so institutionally'.

It is within this context of social structure that the political machine fulfils the basic function of providing avenues of social mobility for the otherwise disadvantaged. Within this context, even the corrupt political machine and the racket 'represent the triumph of amoral intelligence over morally prescribed "failure" when the channels of vertical mobility are closed or narrowed *in a society which places a high premium on economic affluence, [power] and*

social ascent for all its members'. As one sociologist has noted on the basis of several years of close observation in a slum area:

> The sociologist who dismisses racket and political organisations as deviations from desirable standards thereby neglects some of the major elements of slum life. . . . *He does not discover the functions they perform for the members* [of the groupings in the slum]. The Irish and later immigrant peoples have had the greatest difficulty in finding places for themselves in our urban social and economic structure. Does anyone believe that the immigrants and their children could have achieved their present degree of social mobility without gaining control of the political organisation of some of our largest cities? The same is true of the racket organisation. *Politics and the rackets have furnished an important means of social mobility for individuals, who, because of ethnic background and low class position,* are blocked from advancement in the 'respectable' channels.[12]

This, then, represents a third type of function performed for a distinctive subgroup. This function, it may be noted in passing, is fulfilled by the *sheer* existence and operation of the political machine, for it is in the machine itself that these individuals and subgroups find their culturally induced needs more or less satisfied. It refers to the services which the political apparatus provides for its own personnel. But seen in the wider social context we have set forth, it no longer appears as *merely* a means of self-aggrandisement for profit-hungry and power-hungry *individuals*, but as an organised provision for *subgroups* otherwise excluded from or handicapped in the race for 'getting ahead'.

Just as the political machine performs services for 'legitimate' business, so it operates to perform not dissimilar services for 'illegitimate' business: vice, crime and rackets. Once again, the basic sociological role of the machine in this respect can be more fully appreciated only if one temporarily abandons attitudes of moral indignation, to examine in all moral innocence the actual workings of the organisation. In this light, it at once appears that the subgroup of the professional criminal, racketeer or gambler has basic similarities of organisation, demands and operation to the subgroup of the industrialist, man of business or speculator. If there is a Lumber King or an Oil King, there is also a Vice King or a Racket King. If expansive legitimate business organises administrative and financial syndicates to 'rationalise' and to 'integrate' diverse areas of production and business enterprise, so expansive rackets and crime organise syndicates to bring order to the otherwise chaotic areas of production of illicit goods and services. If legitimate business regards the proliferation of small business enterprises as wasteful and inefficient, substituting, for example, the giant chain stores for hundreds of corner groceries, so illegitimate business adopts the same businesslike attitude and syndicates crime and vice.

Finally, and in many respects most important, is the basic similarity, if not near-identity, of the economic role of 'legitimate' business and of 'illegitimate' business. *Both are in some degree concerned with the provision of goods and services for which there is an economic demand.* Morals aside, they are both businesses, industrial and professional enterprises, dispensing goods and services which some people want, for which there is a market in which goods and services are transformed into commodities. And, in a prevalently market society, we should expect appropriate enterprises to arise whenever there is a market demand for certain goods or services.

As is well known, vice, crime and the rackets *are* 'big business'. Consider only that there have been estimated to be about 500,000 professional prostitutes in the United States of 1950, and compare this with the approximately 200,000 physicians and 350,000 professional registered nurses. It is difficult to estimate which have the larger clientele: the professional men and women of medicine or the professional men and women of vice. It is, of course, difficult to estimate the economic assets, income, profits and dividends of illicit gambling in this country and to compare it with the economic assets, income, profits and dividends of, say, the shoe industry, but it is altogether possible that the two industries are about on a par. No precise figures exist on the annual expenditures on illicit narcotics, and it is probable that these are less than the expenditures on candy, but it is also probable that they are larger than the expenditure on books.

It takes but a moment's thought to recognise that, *in strictly economic terms*, there is no relevant difference between the provision of licit and of illicit goods and services. The liquor traffic illustrates this perfectly. It would be peculiar to argue that prior to 1920 (when the 18th Amendment became effective) the provision of liquor constituted an economic service, that from 1920 to 1933 its

production and sale no longer constituted an economic service dispensed in a market, and that from 1934 to the present, it once again took on a serviceable aspect. Or, it would be *economically* (not morally) absurd to suggest that the sale of bootlegged liquor in the dry state of Kansas is less a response to a market demand than the sale of publicly manufactured liquor in the neighbouring wet state of Missouri. Examples of this sort can of course be multiplied many times over. Can it be held that in European countries, with registered and legalised prostitution, the prostitute contributes an economic service, whereas in this country, lacking legal sanction, the prostitute provides no such service? Or that the professional abortionist is in the economic market where he has approved legal status and that he is out of the economic market where he is legally taboo? Or that gambling satisfies a specific demand for entertainment in Nevada, where it constitutes the largest business enterprise of the larger cities in the state, but that it differs essentially in this respect from motion pictures in the neighbouring state of California?[8]

The failure to recognise that these businesses are only *morally* and not *economically* distinguishable from 'legitimate' businesses has led to badly scrambled analysis. Once the economic identity of the two is recognised, we may anticipate that if the political machine performs functions for 'legitimate big business' it will be all the more likely to perform not dissimilar functions for 'illegitimate big business'. And, of course, such is often the case.

The distinctive function of the political machine for their criminal, vice and racket clientele is to enable them to operate in satisfying the economic demands of a large market without undue interference from the government. Just as big business may contribute funds to the political party war-chest to ensure a minimum of governmental interference, so with big rackets and big crime. In both instances, the political machine can, in varying degrees, provide 'protection'. In both instances, many features of the structural context are identical: (1) market demands for goods and services; (2) the operators' concern with maximising gains from their enterprises; (3) the need for partial control of government which might otherwise interfere with these activities of businessmen; (4) the need for an efficient, powerful and centralised agency to provide an effective liaison of 'business' with government.

Without assuming that the foregoing pages exhaust either the range of functions or the range of subgroups served by the political machine, we can at least see that *it presently fulfils some functions for these diverse subgroups which are not adequately fulfilled by culturally approved or more conventional structures.*

Several additional implications of the functional analysis of the political machine can be mentioned here only in passing, although they obviously require to be developed at length. First, the foregoing analysis has direct implications for *social engineering*. It helps explain why the periodic efforts at 'political reform', 'turning the rascals out' and 'cleaning political house' are typically (though not necessarily) short-lived and ineffectual. It exemplifies a basic theorem: *any attempt to eliminate an existing social structure without providing adequate alternative structures for fulfilling the functions previously fulfilled by the abolished organisation is doomed to failure.* (Needless to say, this theorem has much wider bearing than the one instance of the political machine.) When 'political reform' confines itself to the manifest task of 'turning the rascals out', it is engaging in little more than sociological magic. The reform may for a time bring new figures into the political limelight; it may serve the casual social function of reassuring the electorate that the moral virtues remain intact and will ultimately triumph; it may actually effect a turnover in the personnel of the political machine; it may even, for a time, so curb the activities of the machine as to leave unsatisfied the many needs it has previously fulfilled. But, inevitably, unless the reform also involves a 're-forming' of the social and political structure such that the existing needs are satisfied by alternative structures or unless it involves a change which eliminates these needs altogether, the political machine will return to its integral place in the social scheme of things. *To seek social change, without due recognition of the manifest and latent functions performed by the social organisation undergoing change, is to indulge in social ritual rather than social engineering.* The concepts of manifest and latent functions (or their equivalents) are indispensable elements in the theoretic repertoire of the social engineer. In this crucial sense, these concepts are not 'merely' theoretical (in the abusive sense of the term), but are eminently practical. In the deliberate enactment of social change, they can be ignored only at the price of considerably heightening the risk of failure.

A second implication of this analysis of the political machine also has a bearing upon areas wider than the one we have considered.

The paradox has often been noted that the supporters of the political machine include both the 'respectable' business-class elements who are, of course, opposed to the criminal or racketeer and the distinctly 'unrespectable' elements of the underworld. And, at first appearance, this is cited as an instance of very strange bedfellows. The learned judge is not infrequently called upon to sentence the very racketeer beside whom he sat the night before at an informal dinner of the political bigwigs. The district attorney jostles the exonerated convict on his way to the back room where the Boss has called a meeting. The big businessman may complain almost as bitterly as the big racketeer about the 'extortionate' contributions to the party fund demanded by the Boss. Social opposites meet – in the smoke-filled room of the successful politician.

In the light of a functional analysis all this of course no longer seems paradoxical. Since the machine serves both the businessman and the criminal man, the two seemingly antipodal groups intersect. This points to a more general theorem: *the social functions of an organisation help determine the structure (including the recruitment of personnel involved in the structure), just as the structure helps determine the effectiveness with which the functions are fulfilled.* In terms of social status, the business group and the criminal group are indeed poles apart. But status does not fully determine behaviour and the interrelations between groups. Functions modify these relations. Given their distinctive needs, the several subgroups in the large society are 'integrated', whatever their personal desires or intentions, by the centralising structure which serves these several needs. In a phrase with many implications which require further study, *structure affects function and function affects structure.*

CONCLUDING REMARKS

This review of some salient considerations in structural and functional analysis has done little more than indicate some of the principal problems and potentialities of this mode of sociological interpretation. Each of the items codified in the paradigm require sustained theoretic clarification and cumulative empirical research. But it is clear that in functional theory, stripped of those traditional postulates which have fenced it in and often made it little more than a latter-day rationalisation of existing practices, sociology has one beginning of a systematic and empirically relevant mode of analysis. It is hoped that the direction here indicated will suggest the feasibility and the desirability of further codification of functional analysis. In due course each section of the paradigm will be elaborated into a documented, analysed and codified chapter in the history of functional analysis.

Notes

1. These 'explanations' are 'causal' in design. They profess to indicate the social conditions under which political machines come into being. In so far as they are empirically confirmed, these explanations of course add to our knowledge concerning the problem: how is it that political machines operate in certain areas and not in others? How do they manage to continue? *But these causal accounts are not sufficient.* The functional consequences of the machine, as we shall see, go far towards supplementing the causal interpretation.

2. I trust it is superfluous to add that this hypothesis is not 'in support of the political machine'. The question whether the dysfunctions of the machine outweigh its functions, the question whether alternative structures are not available which may fulfil its functions without necessarily entailing its social dysfunctions, still remain to be considered at an appropriate point. We are here concerned with documenting the statement that moral judgements based *entirely* on an appraisal of manifest functions of a social structure are 'unrealistic' in the strict sense, i.e. they do not take into account other actual consequences of that structure, consequences which may provide basic social support for the structure. As will be indicated later, 'social reforms' or 'social engineering' which ignore latent functions do so on pain of suffering acute disappointments and boomerang effects.

3. Again, as with preceding cases, we shall not consider the possible dysfunctions of the political machine.

4. Edward M. Sait, 'Machine Political', *Encyclopaedia of Social Sciences*, vol. 9, p. 659.

5. A. F. Bentley, *The Process of Government* (Chicago, 1908) chap. 2.

6. Herbert Croley, *Progressive Democracy* (New York, 1914) p. 254, cited by Sait, op. cit., 659.

7. Much the same contrast with official welfare policy is found in Harry Hopkins's open-handed and non-political distribution of unemployment relief in New York State under the governorship of Franklin Delano Roosevelt. As Sherwood reports: 'Hopkins was harshly criticised for these irregular activities by the established welfare agencies, which claimed it was "unprofessional conduct" to hand out work tickets without thorough investigation of each applicant, his own or his family's financial resources and probably his religious affiliations. "Harry told the agency to go to hell," said [Hopkins's associate, Dr Jacob A.] Goldberg.' Robert E. Sherwood, *Roosevelt and Hopkins: An Intimate History* (New York, 1948) p. 30.

8. *The Autobiography of Lincoln Steffens* (New York, 1931) pp. 570, 618.

Deriving largely from Steffens, as he says, F. Stuart Chapin sets forth these functions of the political machine with great clarity. See his *Contemporary American Institutions* (New York, 1934) pp. 40–54, 570.

9. Ibid. pp. 572–3. This helps to explain, as Steffens noted after Police Commissioner Theodore Roosevelt, 'the prominence and respectability of the men and women who intercede for crooks' when these have been apprehended in a periodic effort to 'clean up the political machine'. Cf. Steffens, p. 371, and *passim*.

10. See the National Opinion Research Centre survey of evaluation of occupations which firmly documents the general impression that the manual occupations rate very low indeed in the social scale of values, *even among those who are themselves engaged in manual labour*. Consider this latter point in its full implications. In effect, the cultural and social structure exacts the values of pecuniary and power success even among those who find themselves confined to the stigmatised manual occupations. Against this background, consider the powerful motivation for achieving this type of 'success' by any means whatsoever. A garbage-collector who joins with other Americans in the view that the garbage-collector is 'the lowest of the low' occupations can scarcely have a self-image which is pleasing to him; he is in a 'pariah' occupation in the very society where he is assured that 'all who have genuine merit can get ahead'. Add to this, his occasional recognition that 'he didn't have the same chance as others, no matter what they say', and one perceives the enormous psychological pressure upon him for 'evening up the score' by finding some means, whether strictly legal or not, for moving ahead. All this provides the structural and derivatively psychological background for the 'socially induced need' in *some* groups to find some accessible avenue for social mobility.

11. R. Merton, *Social Theory and Social Structure* (Free Press of Glencoe, Illinois, 1957) chap. 4.

12. William F. Whyte, 'Social Organisation in the Slums', *American Sociological Review* (Feb. 1943) vol. 8, pp. 34–9. Thus, the political machine and the racket represent a special case of organisational adjustment to social conditions. It represents, note, an *organisational* adjustment: definite structures arise and operate to reduce somewhat the acute tensions and problems of individuals caught up in the described conflict between the 'cultural accent on success-for-all' and the 'socially structured fact of unequal opportunities for success'. Other types of *individual* 'adjustment' are possible: lone-wolf crime, psychopathological states, rebellion, retreat by abandoning the culturally approved goals, etc. Likewise, other types of *organisational adjustment* sometimes occur; the racket or the political machine are not *alone* available as organised means for meeting this socially induced problem. Participation in revolutionary organisations, for example, can be seen within this context, as an alternative mode of organisational adjustment. All this bears theoretic notice here, since we might otherwise overlook the basic functional concepts of functional substitutes and functional equivalents, which are to be discussed at length in a subsequent publication.

13. Perhaps the most perceptive statement of this view has been made by Hawkins and Waller. 'The prostitute, the pimp, the peddler of dope, the operator of the gambling hall, the vendor of obscene pictures, the bootlegger, the abortionist, all are productive, all produce services or goods which people desire and for which they are willing to pay. It happens that society has put these goods and services under the ban, but people go on producing them and people go on consuming them, and an act of the legislature does not make them any less a part of the economic system.' 'Critical Notes on the Cost of Crime', *Journal of Criminal Law and Criminology* (1936), vol. 26, pp. 679–94.

THE FUNCTIONS OF CONFLICT

Lewis Coser

CONFLICT within a group . . . may help to establish unity or to re-establish unity and cohesion where it has been threatened by hostile and antagonistic feelings among the members. Yet, not *every* type of conflict is likely to benefit group structure, not that conflict can subserve such functions for *all* groups. Whether social conflict is beneficial to internal adaptation or not depends on the type of issues over which it is fought as well as on the type of social structure within which it occurs. However, types of conflict and types of social structure are not independent variables.

Internal social conflicts which concern goals, values or interests that do not contradict the basic assumptions upon which the relationship is founded tend to be positively functional for the social structure. Such conflicts tend to make possible the readjustment of norms and power relations within groups in accordance with the felt needs of its individual members or subgroups.

Internal conflicts in which the contending parties no longer share the basic values upon which the legitimacy of the social system rests threaten to disrupt the structure.

One safeguard against conflict disrupting the consensual basis of the relationship, however, is contained in the social structure itself: it is provided by the institutionalisation and tolerance of conflict. Whether internal conflict promises to be a means of equilibration of social relations or readjustment of rival claims, or whether it threatens to 'tear apart', depends to a large extent on the social structure within which it occurs.

In every type of social structure there are occasions for conflict, since individuals and subgroups are likely to make from time to time rival claims to scarce resources, prestige or power positions. But social structures differ in the way in which they allow ex-

pression to antagonistic claims. Some show more tolerance of conflict than others.

Closely knit groups in which there exists a high frequency of interaction and high personality involvement of the members have a tendency to suppress conflict. While they provide frequent occasions for hostility (since both sentiments of love and hatred are intensified through frequency of interaction), the acting out of such feelings is sensed as a danger to such intimate relationships, and hence there is a tendency to suppress rather than to allow expression of hostile feelings. In close-knit groups, feelings of hostility tend, therefore, to accumulate and hence to intensify. If conflict breaks out in a group that has consistently tried to prevent expression of hostile feelings, it will be particularly intense for two reasons: First, because the conflict does not merely aim at resolving the immediate issue which led to its outbreak; all accumulated grievances which were denied expression previously are apt to emerge at this occasion. Second, because the total personality involvement of the group members makes for mobilisation of all sentiments in the conduct of the struggle.

Hence, the closer the group, the more intense the conflict. Where members participate with their total personality and conflicts are suppressed, the conflict, if it breaks out nevertheless, is likely to threaten the very root of the relationship.

In groups comprising individuals who participate only segmentally, conflict is less likely to be disruptive. Such groups are likely to experience a multiplicity of conflicts. This in itself tends to constitute a check against the breakdown of consensus: the energies of group members are mobilised in many directions and hence will not concentrate on *one* conflict cutting through the group. Moreover, where occasions for hostility are not permitted to accumulate and conflict is allowed to occur wherever a resolution of tension seems to be indicated, such a conflict is likely to remain focused primarily on the condition which led to its outbreak and not to revive blocked hostility; in this way, the conflict is limited to 'the facts of the case'. One may venture to say that multiplicity of conflicts stands in inverse relation to their intensity.

So far we have been dealing with internal social conflict only. At this point we must turn to a consideration of external conflict, for the structure of the group is itself affected by conflicts with other groups in which it engages or which it prepares for. Groups which are engaged in continued struggle tend to lay claim on the total personality involvement of their members so that internal conflict would tend to mobilise all energies and affects of the members. Hence such groups are unlikely to tolerate more than limited departures from the group unity. In such groups there is a tendency to suppress conflict; where it occurs, it leads the group to break up through splits or through forced withdrawal of dissenters.

Groups which are not involved in continued struggle with the outside are less prone to make claims on total personality involvement of the membership and are more likely to exhibit flexibility of structure. The multiple internal conflicts which they tolerate may in turn have an equilibrating and stabilising impact on the structure.

In flexible social structures, multiple conflicts criss-cross each other and thereby prevent basic cleavages along one axis. The multiple group affiliations of individuals make them participate in various group conflicts so that their total personalities are not involved in any single one of them. Thus segmental participation in a multiplicity of conflicts constitutes a balancing mechanism within the structure.

In loosely structured groups and open societies, conflict, which aims at a resolution of tension between antagonists, is likely to have stabilising and integrative functions for the relationship. By permitting immediate and direct expression of rival claims, such social systems are able to readjust their structures by eliminating the sources of dissatisfaction. The multiple conflicts which they experience may serve to eliminate the causes for dissociation and to re-establish unity. These systems avail themselves, through the toleration and institutionalisation of conflict, of an important stabilising mechanism.

In addition, conflict within a group frequently helps to revitalise existent norms; or it contributes to the emergence of new norms. In this sense, social conflict is a mechanism for adjustment of norms adequate to new conditions. A flexible society benefits from conflict because such behaviour, by helping to create and modify norms, assures its continuance under changed conditions. Such mechanism for readjustment of norms is hardly available to rigid systems: by suppressing conflict, the latter smother a useful warning signal, thereby maximising the danger of catastrophic break-down.

Internal conflict can also serve as a means for ascertaining the

relative strength of antagonistic interests within the structure, and in this way constitutes a mechanism for the maintenance or continual readjustment of the balance of power. Since the outbreak of the conflict indicates a rejection of a previous accommodation between parties, once the respective power of the contenders has been ascertained through conflict, a new equilibrium can be established and the relationship can proceed on this new basis. Consequently, a social structure in which there is room for conflict disposes of an important means for avoiding or redressing conditions of disequilibrium by modifying the terms of power relations.

Conflicts with some produce associations or coalitions with others. Conflicts through such associations or coalitions, by providing a bond between the members, help to reduce social isolation or to unite individuals and groups otherwise unrelated or antagonistic to each other. A social structure in which there can exist a multiplicity of conflicts contains a mechanism for bringing together otherwise isolated, apathetic or mutually hostile parties and for taking them into the field of public social activities. Moreover, such a structure fosters a multiplicity of associations and coalitions whose diverse purposes criss-cross each other, we recall, thereby preventing alliances along one major line of cleavage.

Once groups and associations have been formed through conflict with other groups, such conflict may further serve to maintain boundary lines between them and the surrounding social environment. In this way, social conflict helps to structure the larger social environment by assigning position to the various subgroups within the system and by helping to define the power relations between them.

Not all social systems in which individuals participate segmentally allow the free expression of antagonistic claims. Social systems tolerate or institutionalise conflict to different degrees. There is no society in which any and every antagonistic claim is allowed immediate expression. Societies dispose of mechanisms to channel discontent and hostility while keeping intact the relationship within which antagonism arises. Such mechanisms frequently operate through 'safety-valve' institutions which provide substitute objects upon which to displace hostile sentiments as well as means of abreaction of aggressive tendencies.

Safety-valve institutions may serve to maintain both the social structure and the individual's security system, but they are incompletely functional for both of them. They prevent modification of relationships to meet changing conditions and hence the satisfaction they afford the individual can be only partially or momentarily adjustive. The hypothesis has been suggested that the need for safety-valve institutions increases with the rigidity of the social structure, i.e. with the degree to which it disallows direct expression of antagonistic claims.

Safety-valve institutions lead to a displacement of goal in the actor: he need no longer aim at reaching a solution of the unsatisfactory situation, but merely at releasing the tension which arose from it. Where safety-valve institutions provide substitute objects for the displacement of hostility, the conflict itself is channelled away from the original unsatisfactory relationship into one in which the actor's goal is no longer the attainment of specific results, but the release of tension.

This affords us a criterion for distinguishing between realistic and non-realistic conflict.

Social conflicts that arise from frustrations of specific demands within a relationship and from estimates of gains of the participants, and that are directed at the presumed frustrating object, can be called realistic conflicts. In so far as they are means towards specific results, they can be replaced by alternative modes of interaction with the contending party if such alternatives seem to be more adequate for realising the end in view.

Non-realistic conflicts, on the other hand, are not occasioned by the rival ends of the antagonists, but by the need for tension release of one or both of them. In this case the conflict is not oriented towards the attainment of specific results. In so far as unrealistic conflict is an end in itself, in so far as it affords only tension release, the chosen antagonist can be substituted for by any other 'suitable' target.

In realistic conflict, there exist functional alternatives with regard to the means of carrying out the conflict, as well as with regard to accomplishing desired results short of conflict; in non-realistic conflict, on the other hand, there exist only functional alternatives in the choice of antagonists.

Our hypothesis, that the need for safety-valve institutions increases with the rigidity of the social system, may be extended to suggest that unrealistic conflict may be expected to occur as a consequence of rigidity present in the social structure.

Our discussion of the distinction between types of conflict, and between types of social structures, leads us to conclude that conflict tends to be dysfunctional for a social structure in which there is no or insufficient toleration and institutionalisation of conflict. The intensity of a conflict which threatens to 'tear apart', which attacks the consensual basis of a social system, is related to the rigidity of the structure. What threatens the equilibrium of such a structure is not conflict as such, but the rigidity itself which permits hostilities to accumulate and to be channelled along one major line of cleavage once they break out in conflict.

In this chapter we are concerned with the behaviour of individuals in groups.

The first paper included is a well-known classic: 'Effects of Group Pressure upon the Modification and Distortion of Judgements' by Solomon Asch. Asch investigates the effects of majority opinions on the judgements made by individuals. His experimental evidence shows that a number of factors interact. Whether an individual resists group pressures or yields to them depends on his own personality, the structure of the group and the characteristics of the stimulus situation.

The second article is also a classic: 'Attitude Development as a Function of Reference Groups: The Bennington Study' by Theodore M. Newcomb. Newcomb studied, not in the laboratory, but in the 'real-life' situation of a residential college, the changing attitudes of students towards public issues. In order to explain why some students change and why others do not change their opinions in their new environment, he distinguishes between 'membership' groups and 'reference' groups, and he discusses the implications for understanding attitude change which arise from this distinction. A membership group (that is, one in which one is physically present) only affects one's attitudes if it is also taken as a positive reference group for these particular attitudes. An individual, therefore, does not simply adopt the current and approved views of the group he finds himself with; he may, for instance, continue to refer his outlook to a previous membership group and this group may thus remain his point of reference for, say, political attitudes.

We also include a disturbing paper by Stanley Milgram, 'Some Conditions of Obedience and Disobedience to Authority'. He discusses here his researches at Yale University into the conditions under which a 'subject' will inflict pain on a 'victim' or refuse to do so at the command of the experimenter. Of particular relevance in the present context are the experimental variations in which the 'subject' is in a group which (by prior arrangement with the experimenter) defies or supports the experimenter's authority. What is perhaps particularly disturbing is the speed with which the experimenter, the 'subject' and the experimenter's 'stooges' can form a group from which the 'victim' is excluded, and to whom, as an outsider, ordinary considerations of decency do not apply.

EFFECTS OF GROUP PRESSURE UPON THE MODIFICATION AND DISTORTION OF JUDGEMENTS

S. E. Asch

We shall here describe in summary form the conception and first findings of a programme of investigation into the conditions of independence and submission to group pressure.

Our immediate object was to study the social and personal conditions that induce individuals to resist or to yield to group pressures when the latter are perceived to be *contrary to fact*. The issues which this problem raises are of obvious consequence for society; it can be of decisive importance whether or not a group will, under certain conditions, submit to existing pressures. Equally direct are the consequences for individuals and our understanding of them, since it is a decisive fact about a person whether he possesses the freedom to act independently, or whether he characteristically submits to group pressures.

The problem under investigation requires the direct observation of certain basic processes in the interaction between individuals, and between individuals and groups. To clarify these seems necessary if we are to make fundamental advances in the understanding of the formation and reorganisation of attitudes, of the functioning of public opinion, and of the operation of propaganda. Today we do not possess an adequate theory of these central psycho-social processes. Empirical investigation has been predominantly controlled by general propositions concerning group influence which have as a rule been assumed but not tested. With few exceptions investigation has relied upon descriptive formulations concerning the operation of suggestion and prestige, the inadequacy of which is becoming increasingly obvious, and upon schematic applications of stimulus-response theory.

Basic to the current approach has been the axiom that group pressures characteristically induce psychological changes *arbitrarily*, in far-reaching disregard of the material properties of the

given conditions. This mode of thinking has almost exclusively stressed the slavish submission of individuals to group forces, has neglected to inquire into their possibilities for independence and for productive relations with the human environment, and has virtually denied the capacity of men under certain conditions to rise above group passion and prejudice. It was our aim to contribute to a clarification of these questions, important both for theory and for their human implications, by means of direct observation of the effects of groups upon the decisions and evaluations of individuals.

THE EXPERIMENT AND FIRST RESULTS

To this end we developed an experimental technique which has served as the basis for the present series of studies. We employed the procedure of placing an individual in a relation of radical conflict with all the other members of a group, of measuring its effect upon him in quantitative terms and of describing its psychological consequences. A group of eight individuals was instructed to judge a series of simple, clearly structured perceptual relations – to match the length of a given line with one of three unequal lines. Each member of the group announced his judgements publicly. In the midst of this monotonous 'test' one individual found himself suddenly contradicted by the entire group, and this contradiction was repeated again and again in the course of the experiment. The group in question had, with the expection of one member, previously met with the experimenter and received instructions to respond at certain points with wrong – and unanimous – judgements. The errors of the majority were large (ranging between $\frac{1}{2}$ in. and $1\frac{3}{4}$ in.) and of an order not encountered under control conditions. The outstanding person – the critical subject – whom we had placed in the position of a *minority of one* in the midst of a *unanimous majority* – was the object of investigation. He faced, possibly for the first time in his life, a situation in which a group unanimously contradicted the evidence of his senses.

This procedure was the starting point of the investigation and the point of departure for the study of further problems. Its main features were the following: (1) The critical subject was submitted to two contradictory and irreconcilable forces – the evidence of his own experience of a clearly perceived relation, and the unanimous evidence of a group of equals. (2) Both forces were part of the immediate situation; the majority was concretely present, surrounding the subject physically. (3) The critical subject, who was requested together with all others to state his judgements publicly, was obliged to declare himself and to take a definite stand *vis-à-vis* the group. (4) The situation possessed a self-contained character. The critical subject could not avoid or evade the dilemma by reference to conditions external to the experimental situation. (It may be mentioned at this point that the forces generated by the given conditions acted so quickly upon the critical subjects that instances of suspicion were infrequent.)

The technique employed permitted a simple quantitative measure of the 'majority effect' in terms of the frequency of errors in the direction of the distorted estimates of the majority. At the same time we were concerned to obtain evidence of the ways in which the subjects perceived the group, to establish whether they became doubtful, whether they were tempted to join the majority. Most important, it was our object to establish the grounds of the subject's independence or yielding – whether, for example, the yielding subject was aware of the effect of the majority upon him, whether he abandoned his judgement deliberately or compulsively. To this end we constructed a comprehensive set of questions which served as the basis of an individual interview immediately following the experimental period. Towards the conclusion of the interview each subject was informed fully of the purpose of the experiment, of his role and of that of the majority. The reactions to the disclosure of the purpose of the experiment became in fact an integral part of the procedure. The information derived from the interview became an indispensable source of evidence and insight into the psychological structure of the experimental situation, and in particular, of the nature of the individual differences. It should be added that it is not justified or advisable to allow the subject to leave without giving him a full explanation of the experimental conditions. The experimenter has a responsibility to the subject to clarify his doubts and to state the reasons for placing him in the experimental situation. When this is done most subjects react with interest, and some express gratification at having lived through a striking situation which has some bearing on them personally and on wider human issues.

Both the members of the majority and the critical subjects were

male college students. We shall report the results for a total of fifty critical subjects in this experiment. In Table 31.1 we summarise the successive comparison trials and the majority estimates. The reader will note that on certain trials the majority responded correctly; these were the 'neutral' trials. There were twelve critical trials on which the responses of the majority responded incorrectly.

TABLE 31.1

LENGTHS OF STANDARD AND COMPARISON LINES

Trial	*Length of standard line (in inches)*	*Comparison lines (in inches)*			*Correct response*	*Group response*	*Majority error (in inches)*
		1	*2*	*3*			
1	10	$8\frac{3}{4}$	10	8	2	2	—
2	2	2	1	$1\frac{1}{2}$	1	1	—
3	3	$3\frac{3}{4}$	$4\frac{1}{4}$	3	3	1*	$+\frac{3}{4}$
4	5	5	5	$6\frac{1}{2}$	1	2*	−1·0
5	4	3	5	4	3	3	—
6	3	$3\frac{3}{4}$	$4\frac{1}{4}$	3	3	2*	$+1\frac{1}{4}$
7	8	$6\frac{1}{4}$	8	$6\frac{3}{4}$	2	3*	$-1\frac{1}{4}$
8	5	5	4	$6\frac{1}{2}$	1	3*	$+1\frac{1}{2}$
9	8	$6\frac{1}{4}$	8	$6\frac{3}{4}$	2	1*	$-1\frac{3}{4}$
10	10	$8\frac{3}{4}$	10	8	2	2	—
11	2	2	1	$1\frac{1}{2}$	1	1	—
12	3	$3\frac{3}{4}$	$4\frac{1}{4}$	3	3	1*	$+\frac{3}{4}$
13	5	5	4	$6\frac{1}{2}$	1	2*	−1·0
14	4	3	5	4	3	3	—
15	3	$3\frac{3}{4}$	$4\frac{1}{4}$	3	3	2*	$+1\frac{1}{4}$
16	8	$6\frac{1}{4}$	8	$6\frac{3}{4}$	2	3*	$-1\frac{1}{4}$
17	5	5	4	$6\frac{1}{2}$	1	3*	$+1\frac{1}{2}$
18	8	$6\frac{1}{4}$	8	$6\frac{3}{4}$	2	1*	$-1\frac{3}{4}$

*Starred figures designate the erroneous estimates by the majority.

The quantitative results are clear and unambiguous.

1. There was a marked movement towards the majority. One-third of all the estimates in the critical group were errors identical with or in the direction of the distorted estimates of the majority. The significance of this finding becomes clear in the light of the virtual absence of errors in the control group, the members of which recorded their estimates in writing. The relevant data of the critical and control groups are summarised in Table 31.2.

2. At the same time the effect of the majority was far from complete. The preponderance of estimates in the critical group (68 per cent) was correct despite the pressure of the majority.

3. We found evidence of extreme individual differences. There were in the critical group subjects who remained independent without exception, and there were those who went nearly all the time with the majority. (The maximum possible number of errors was 12, while the actual range of errors was 0–11.) One-fourth of the critical subjects was completely independent; at the other extreme, one-third of the group displaced the estimates towards the majority in one-half or more of the trials.

The differences between the critical subjects in their reactions to the given conditions were equally striking. There were subjects who remained completely confident throughout. At the other extreme were those who became disoriented, doubt-ridden and experienced a powerful impulse not to appear different from the majority.

TABLE 31.2

DISTRIBUTION OF ERRORS IN EXPERIMENTAL AND CONTROL GROUPS

Number of critical errors	*Critical group** (N = 50) *F*	*Control group* (N = 37) *F*
0	13	35
1	4	1
2	5	1
3	6	
4	3	
5	4	
6	1	
7	2	
8	5	
9	3	
10	3	
11	1	
12	0	
Total	50	37
Mean	3·84	0·08

*All errors in the critical group were in the direction of the majority estimates.

For purposes of illustration we include a brief description of one independent and one yielding subject.

Independent

After a few trials he appeared puzzled, hesitant. He announced all disagreeing answers in the form of 'Three, sir; two, sir'; not so with the unanimous answers on the neutral trials. At Trial 4 he answered immediately after the first member of the group, shook his head, blinked, and whispered to his neighbour: 'Can't help it, that's one.' His later answers came in a whispered voice, accompanied by a deprecating smile. At one point he grinned embarrassedly, and whispered explosively to his neighbour: 'I always disagree – darn it!' During the questioning, this subject's constant refrain was: 'I called them as I saw them, sir.' He insisted that his estimates were right without, however, committing himself as to whether the others were wrong, remarking that 'that's the way I see them and that's the way they see them'. If he had to make a practical decision under similar circumstances, he declared, 'I would follow my own view, though part of my reason would tell me that I might be wrong.' Immediately following the experiment the majority engaged this subject in a brief discussion. When they pressed him to say whether the entire group was wrong and he alone right, he turned upon them defiantly, exclaiming: 'You're *probably* right, but you *may* be wrong!' To the disclosure of the experiment this subject reacted with the statement that he felt 'exultant and relieved', adding, 'I do not deny that at times I had the feeling: "to heck with it, I'll go along with the rest." '

Yielding

This subject went with the majority in eleven out of twelve trials. He appeared nervous and somewhat confused, but he did not attempt to evade discussion; on the contrary, he was helpful and tried to answer to the best of his ability. He opened the discussion with the statement: 'If I'd been first I probably would have responded differently'; this was his way of stating that he had adopted the majority estimates. The primary factor in his case was loss of confidence. He perceived the majority as a decided group, acting without hesitation: 'If they had been doubtful I probably would have changed, but they answered with such confidence.' Certain of his errors, he explained, were due to the doubtful nature of the comparisons; in such instances he went with the majority. When the object of the experiment was explained, the subject volunteered: 'I suspected about the middle – but tried to push it out of my mind.' It is of interest that his suspicion did not restore his confidence or diminish the power of the majority. Equally striking is his report that he assumed the experiment to involve an 'illusion' to which the others, but not he, were subject. This assumption too did not help to free him; on the contrary, he acted as if his divergence from the majority was a sign of defect. The principal impression this subject produced was of one so caught up by immediate difficulties that he lost clear reasons for his actions, and could make no reasonable decisions.

A FIRST ANALYSIS OF INDIVIDUAL DIFFERENCES

On the basis of the interview data described earlier, we undertook to differentiate and describe the major forms of reaction to the experimental situation, which we shall now briefly summarise.

Among the *independent* subjects we distinguished the following main categories:

1. Independence based on *confidence* in one's perception and experience. The most striking characteristic of these subjects is the vigour with which they withstand the group opposition. Though they are sensitive to the group, and experience the conflict, they show a resilience in coping with it, which is expressed in their continuing reliance on their perception and the effectiveness with which they shake off the oppressive group opposition.

2. Quite different are those subjects who are independent and *withdrawn*. These do not react in a spontaneously emotional way, but rather on the basis of explicit principles concerning the necessity of being an individual.

3. A third group of independent subjects manifests considerable tension and doubt, but adhere to their judgement on the basis of a felt necessity to deal adequately with the task.

The following were the main categories of reaction among the *yielding* subjects, or those who went with the majority during one-half or more of the trials.

1. *Distortion of perception* under the stress of group pressure. In this category belong a very few subjects who yield completely, but are not aware that their estimates have been displaced or distorted by the majority. These subjects report that they came to perceive the majority estimates as correct.

2. *Distortion of judgement*. Most submitting subjects belong to this category. The factor of greatest importance in this group is a decision the subjects reach that their perceptions are inaccurate, and that those of the majority are correct. These subjects suffer from primary doubt and lack of confidence; on this basis they feel a strong tendency to join the majority.

3. *Distortion of action*. The subjects in this group do not suffer a modification of perception nor do they conclude that they are wrong. They yield because of an overmastering need not to appear different from or inferior to others, because of an inability to tolerate the appearance of defectiveness in the eyes of the group. These subjects suppress their observations and voice the majority position with awareness of what they are doing.

The results are sufficient to establish that independence and yielding are not psychologically homogeneous, that submission to group pressure and freedom from pressure can be the result of different psychological conditions. It should also be noted that the categories described above, being based exclusively on the subjects' reactions to the experimental conditions, are descriptive, not presuming to explain why a given individual responded in one way rather than another. The further exploration of the basis for the individual differences is a separate task.

EXPERIMENTAL VARIATIONS

The results described are clearly a joint function of two broadly different sets of conditions. They are determined first by the specific external conditions, by the particular character of the relation between social evidence and one's own experience. Second, the presence of pronounced individual differences points to the important role of personal factors, or factors connected with the individual's character structure. We reasoned that there are group conditions which would produce independence in all subjects, and that there probably are group conditions which would induce intensified yielding in many, though not in all. Secondly, we deemed it reasonable to assume that behaviour under the experimental social pressure is significantly related to certain characteristics of the individual. The present account will be limited to the effect of the surrounding conditions upon independence and submission. To this end we followed the procedure of experimental variation, systematically altering the quality of social evidence by means of systematic variation of the group conditions and of the task.

The effect of non-unanimous majorities

Evidence obtained from the basic experiment suggested that the condition of being exposed *alone* to the opposition of a 'compact majority' may have played a decisive role in determining the course and strength of the effects observed. Accordingly we undertook to investigate in a series of successive variations the effects of *non-unanimous* majorities. The technical problem of altering the uniformity of a majority is, in terms of our procedure, relatively simple. In most instances we merely directed one or more members

of the instructed group to deviate from the majority in prescribed ways. It is obvious that we cannot hope to compare the performance of the same individual in two situations on the assumption that they remain independent of one another; at best we can investigate the effect of an earlier upon a later experimental condition. The comparison of different experimental situations therefore requires the use of different but comparable groups of critical subjects. This is the procedure we have followed. In the variations to be described we have maintained the conditions of the basic experiment (e.g. the sex of the subjects, the size of the majority, the content of the task and so on) save for the specific factor that was varied. The following were some of the variations studied:

1. *The presence of a 'true partner'.* (*a*) In the midst of the majority were *two* naïve, critical subjects. The subjects were separated spatially, being seated in the fourth and eighth positions, respectively. Each therefore heard his judgements confirmed by one other person (provided the other person remained independent), one prior to, the other after announcing his own judgement. In addition, each experienced a break in the unanimity of the majority. There were six pairs of critical subjects. (*b*) In a further variation the 'partner' to the critical subject was a member of the group who had been instructed to respond correctly throughout. This procedure permits the exact control of the partner's responses. The partner was always seated in the fourth position; he therefore announced his estimates in each case before the critical subject.

The results clearly demonstrate that a disturbance of the unanimity of the majority markedly increased the independence of the critical subjects. The frequency of pro-majority errors dropped to 10·4 per cent of the total number of estimates in variation (*a*), and to 5·5 per cent in variation (*b*). These results are to be compared with the frequency of yielding to the unanimous majorities in the basic experiment, which was 32 per cent of the total number of estimates. It is clear that the presence in the field of *one other* individual who responded correctly was sufficient to deplete the power of the majority, and in some cases to destroy it. This finding is all the more striking in the light of other variations which demonstrate the effect of even small minorities provided they are unanimous. Indeed, we have been able to show that a unanimous majority of 3 is, under the given conditions, far more effective than a majority of 8 containing 1 dissenter. That critical subjects will under these conditions free themselves of a majority of 7 and join forces with one other person in the minority is, we believe, a result significant for theory. It points to a fundamental psychological difference between the condition of being alone and having a minimum of human support. It further demonstrates that the effects obtained are not the result of a summation of influences proceeding from each member of the group; it is necessary to conceive the results as being relationally determined.

2. *Withdrawal of a 'true partner'.* What will be the effect of providing the critical subject with a partner who responds correctly and then withdrawing him? The critical subject started with a partner who responded correctly. The partner was a member of the majority who had been instructed to respond correctly and to 'desert' to the majority in the middle of the experiment. This procedure permits the observation of the same subject in the course of the transition from one condition to another. The withdrawal of the partner produced a powerful and unexpected result. We had assumed that the critical subject, having gone through the experience of opposing the majority with a minimum of support, would maintain his independence when alone. Contrary to this expectation, we found that the experience of having had and then lost a partner restored the majority effect to its full force, the proportion of errors rising to 28·5 per cent of all judgements, in contrast to the preceding level of 5·5 per cent. Further experimentation is needed to establish whether the critical subjects were responding to the sheer fact of being alone, or to the fact that the partner abandoned them.

3. *Late arrival of a 'true partner'.* The critical subject started as a minority of 1 in the midst of a unanimous majority. Towards the conclusion of the experiment one member of the majority 'broke' away and began announcing correct estimates. This procedure, which reverses the order of conditions of the preceding experiment, permits the observation of the transition from being alone to being a member of a pair against a majority.

It is obvious that those critical subjects who were independent when alone would continue to be so when joined by a partner. The variation is therefore of significance primarily for those subjects who yielded during the first phase of the experiment. The appearance of the late partner exerts a freeing effect, reducing the level of yielding to 8·7 per cent. Those who had previously yielded also became markedly more independent, but not completely so, continuing to yield more than previously independent subjects. The reports of the subjects do not cast much light on the factors responsible for the result. It is our impression that some subjects, having once committed themselves to yielding, find it difficult to change their direction completely. To do so is tantamount to a public admission that they had not acted rightly. They therefore follow to an extent the precarious course they had chosen in order to maintain an outward semblance of consistency and conviction.

4. *The presence of a 'compromise partner'*. The majority was consistently extremist, always matching the standard with the most unequal line. One instructed subject (who, as in the other variations, preceded the critical subject) also responded incorrectly, but his estimates were always intermediate between the truth and the majority position. The critical subject therefore faced an extremist majority whose unanimity was broken by one more moderately erring person. Under these conditions the frequency of errors was reduced but not significantly. However, the lack of unanimity determined in a strikingly consistent way the *direction* of the errors. The preponderance of the errors, 75·7 per cent of the total, was moderate, whereas in a parallel experiment in which the majority was unanimously extremist (i.e. with the 'compromise' partner excluded), the incidence of moderate errors was 42 per cent of the total. As might be expected, in a unanimously moderate majority, the errors of the critical subjects were without exception moderate.

The role of majority size

To gain further understanding of the majority effect, we varied the size of the majority in several different variations. The majorities, which were in each case unanimous, consisted of 2, 3, 4, 8 and 10–15 persons, respectively. In addition, we studied the limiting case in which the critical subject was opposed by one instructed subject. Table 31.3 contains the mean and the range of errors under each condition.

With the opposition reduced to 1, the majority effect all but disappeared. When the opposition proceeded from a group of 2, it produced a measurable though small distortion, the errors being 12·8 per cent of the total number of estimates. The effect appeared in full force with a majority of 3. Larger majorities did not produce effects greater than a majority of 3.

The effect of a majority is often silent, revealing little of its operation to the subject, and often hiding it from the experimenter. To examine the range of effects it is capable of inducing, decisive variations of conditions are necessary. An indication of one effect is furnished by the following variation in which the conditions of the basic experiment were simply reversed. Here the majority, consisting of a group of 16, was naïve; in the midst of it we placed a single individual who responded wrongly according to instructions. Under these conditions the members of the naïve majority reacted to the lone dissenter with amusement. Contagious laughter spread through the group at the droll minority of 1. Of significance is the fact that the members lacked awareness that they drew their strength from the majority, and that their reactions would change radically if they faced the dissenter individually. These observations demonstrate the role of social support as a source of power and stability, in contrast to the preceding investigations which stressed the effects of social opposition. Both aspects must be explicitly considered in a unified formulation of the effects of group conditions on the formation and change of judgements.

TABLE 31.3

ERRORS OF CRITICAL SUBJECTS WITH UNANIMOUS MAJORITIES OF DIFFERENT SIZE

Size of majority	*Control*	*1*	*2*	*3*	*4*	*8*	*10–15*
N	37	10	15	10	10	50	12
Mean number of errors	0·08	0·33	1·53	4·0	4·20	3·84	3·75
Range of errors	0–2	0–1	0–5	1–12	0–11	0–11	0–10

The role of the stimulus-situation

It is obviously not possible to divorce the quality and course of the group forces which act upon the individual from the specific stimulus-conditions. Of necessity the structure of the situation moulds the group forces and determines their direction as well as their strength. Indeed, this was the reason that we took pains in the investigations described above to centre the issue between the individual and the group around an elementary matter of fact. And there can be no doubt that the resulting reactions were directly a function of the contradiction between the observed relations and the majority position. These general considerations are sufficient to establish the need to vary the stimulus-conditions and to enforce their effect on the resulting group forces.

Accordingly we have studied the effect of increasing and decreasing the discrepancy between the correct relation and the position of the majority, going beyond the basic experiment which contained discrepancies of a relatively moderate order. Our technique permits the easy variation of this factor, since we can vary at will the deviation of the majority from the correct relation. At this point we can only summarise the trend of the results which is entirely clear. The degree of independence increases with the distance of the majority from correctness. However, even glaring discrepancies (of the order of 3–6 in.) did not produce independence in all. While independence increases with the magnitude of contradiction, a certain proportion of individuals continues to yield under extreme conditions.

We have also varied systematically the structural clarity of the task, employing judgements based on mental standards. In agreement with other investigators, we find that the majority effect grows stronger as the situation diminishes in clarity. Concurrently, however, the disturbance of the subjects and the conflict-quality of the situation decrease markedly. We consider it of significance that the majority achieves its most pronounced effect when it acts most painlessly.

SUMMARY

We have investigated the effects upon individuals of majority opinions when the latter were seen to be in a direction contrary to fact. By means of a simple technique we produced a radical divergence between a majority and a minority, and observed the ways in which individuals coped with the resulting difficulty. Despite the stress of the given conditions, a substantial proportion of individuals retained their independence throughout. At the same time a substantial minority yielded, modifying their judgements in accordance with the majority. Independence and yielding are a joint function of the following major factors: (1) The character of the stimulus situation. Variations in structural clarity have a decisive effect: with diminishing clarity of the stimulus-conditions the majority effect increases. (2) The character of the group forces. Individuals are highly sensitive to the structural qualities of group opposition. In particular, we demonstrated the great importance of the factor of unanimity. Also, the majority effect is a function of the size of group opposition. (3) The character of the individual. There were wide and, indeed, striking differences among individuals within the same experimental situation.

ATTITUDE DEVELOPMENT AS A FUNCTION OF REFERENCE GROUPS: THE BENNINGTON STUDY

Theodore M. Newcomb

MEMBERSHIP in established groups usually involves the taking on of whole patterns of interrelated behaviour and attitudes. This was one of the hypotheses pursued in the study which is reported here in part. The group selected for study consisted of the entire student body at Bennington College – more than 600 individuals – between the years 1935 and 1939. One of the problems to be investigated was that of the manner in which the patterning of behaviour and attitudes varied with different degrees of assimilation into the community.

Not all of the attitudes and behaviours that are likely to be taken on by new members, as they become absorbed into a community, can be investigated in a single study. A single, though rather inclusive, area of adaptation to the college community was therefore selected for special study, namely, *attitudes towards public affairs*. There were two reasons for this selection: (1) methods of attitude measurement were readily available; and (2) there was an

unusually high degree of concern, in this community at this time, over a rather wide range of public issues. This latter fact resulted partly from the fact that the college opened its doors during the darkest days of the depression of the 1930s, and its formative period occurred in the period of social change characterised by the phrase 'the New Deal'. This was also the period of gathering war clouds in Europe. Underlying both of these circumstances, however, was the conviction on the part of the faculty that one of the foremost duties of the college was to acquaint its somewhat oversheltered students with the nature of their contemporary social world.

In a membership group in which certain attitudes are approved (i.e. held by majorities, and conspicuously so by leaders), individuals acquire the approved attitudes to the extent that the membership group (particularly as symbolised by leaders and dominant subgroups) serves as a positive point of reference. The findings of the Bennington study seem to be better understood in terms of this thesis than any other. The distinction between membership group and reference group is a crucial one, in fact, although the original report did not make explicit use of it.

The above statement does not imply that no reference groups other than the membership group are involved in attitude formation; as we shall see, this is distinctly not the case. Neither does it imply that the use of the membership group as reference group necessarily results in adoption of the approved attitudes. It may also result in their rejection; hence the word *positive* in the initial statement. It is precisely these variations in degree and manner of relationship between reference group and membership group which must be known in order to explain individual variations in attitude formation, as reported in this study.

The essential facts about the Bennington membership group are as follows: (1) It was small enough (about 250 women students) so that data could be obtained from every member. (2) It was in most respects self-sufficient; college facilities provided not only the necessities of living and studying, but also a co-operative store, post office and Western Union office, beauty parlour, gasoline station, and a wide range of recreational opportunities. The average student visited the four-mile-distant village once a week and spent one weekend a month away from the college. (3) It was self-conscious and enthusiastic, in large part because it was new the study was begun during the first year in which there was a senior class) and because of the novelty and attractiveness of the college's educational plan. (4) It was unusually active and concerned about public issues, largely because the faculty felt that its educational duties included the familiarising of an oversheltered student body with the implications of a depression-torn America and a war-threatened world. (5) It was relatively homogeneous in respect to home background; tuition was very high, and the large majority of students came from urban, economically privileged families whose social attitudes were conservative.

Most individuals in this total membership group went through rather marked changes in attitudes towards public issues, as noted below. In most cases the total membership group served as the reference group for the changing attitudes. But some individuals changed little or not at all in attitudes during the four years of the study; attitude persistence was in some of these cases a function of the membership group as reference group and in some cases it was not. Among those who did change, moreover, the total membership group sometimes served as reference group but sometimes it did not. An over-simple theory of 'assimilation into the community' thus leaves out of account some of those whose attitudes did and some of those whose attitudes did not change; they remain unexplained exceptions. A theory which traces the impact of other reference groups as well as the effect of the membership group seems to account for all cases without exception.

The general trend of attitude change for the total group is from freshman conservatism to senior non-conservatism (as the term was commonly applied to the issues towards which attitudes were measured). During the 1936 presidential election, for example, 62 per cent of the freshmen and only 14 per cent of the juniors and seniors 'voted' for the Republican candidate, 29 per cent of freshmen and 54 per cent of juniors and seniors for Roosevelt, and 9 per cent of freshmen as compared with 30 per cent of juniors and seniors for the Socialist or Communist candidates. Attitudes towards nine specific issues were measured during the four years of the study, and seniors were less conservative in all of them than freshmen; six of the nine differences are statistically reliable. These differences are best shown by a Likert-type scale labelled Political and Economic Progressivism (P.E.P.) which dealt with such

issues as unemployment, public relief and the rights of organised labour, which were made prominent by the New Deal. Its odd–even reliability was about 0·9, and it was given once or more during each of the four years of the study to virtually all students. The critical ratios of the differences between freshmen and juniors–seniors in four successive years ranged between 3·9 and 6·5; the difference between the average freshman and senior scores of 44 individuals (the entire class that graduated in 1939) gives a critical ratio of 4·3.

As might be anticipated in such a community, *individual prestige was associated with non-conservatism.* Frequency of choice as one of five students 'most worthy to represent the College' at an intercollegiate gathering was used as a measure of prestige. Nominations were submitted in sealed envelopes by 99 per cent of all students in two successive years, with almost identical results. The non-conservatism of those with high prestige is not merely the result of the fact that juniors and seniors are characterised by both high prestige and non-conservatism; in each class those who have most prestige are least conservative. For example, ten freshmen receiving 2 to 4 choices had an average P.E.P. score of 64·6 as compared with 72·8 for freshmen not chosen at all (high scores are conservative); eight sophomores chosen 12 or more times had an average score of 63·6 as compared with 71·3 for those not chosen; the mean P.E.P. score of five juniors and seniors chosen 40 or more times was 50·4 and of the fifteen chosen 12 to 39 times, 57·6, as compared with 69·0 for those not chosen. In each class, those intermediate in prestige are also intermediate in average P.E.P. score.

Such were the attitudinal characteristics of the total membership group, expressed in terms of average scores. Some individuals, however, showed these characteristics in heightened form and others failed to show them at all. An examination of the various reference groups in relation to which attitude change did or did not occur, and of the ways in which they were brought to bear, will account for a large part of such attitude variance.

Information concerning reference groups was obtained both directly, from the subjects themselves, and indirectly, from other students and from teachers. Chief among the indirect procedures was the obtaining of indexes of 'community citizenship' by a guess-who technique. Each of twenty-four students, carefully selected to represent every cross-section and grouping of importance within the community, named three individuals from each of three classes who were reputedly most extreme in each of twenty-eight characteristics related to community citizenship. The relationship between reputation for community identification and non-conservatism is a close one, in spite of the fact that no reference was made to the latter characteristic when the judges made their ratings. A reputation index was computed, based upon the frequency with which individuals were named in five items dealing with identification with the community, minus the number of times they were named in five other items dealing with negative community attitude. Examples of the former items are: 'absorbed in college community affairs', and 'influenced by community expectations regarding codes, standards, etc.'; examples of the latter are: 'indifferent to activities of student committees', and 'resistant to community expectations regarding codes, standards, etc.' The mean senior P.E.P. score of fifteen individuals whose index was +15 or more was 54·4; of sixty-three whose index was +4 to −4, 65·3; and of ten whose index was −15 or less, 68·2.

To have the reputation of identifying oneself with the community is not the same thing, however, as to identify the community as a reference group for a specific purpose – e.g. in this case, as a point of reference for attitudes towards public issues. In short, the reputation index is informative as to degree and direction of tendency to use the total membership group as a *general* reference group, but not necessarily as a group to which social attitudes are referred. For this purpose information was obtained directly from students.

Informal investigation had shown that whereas most students were aware of the marked freshman-to-senior trend away from conservatism, a few (particularly among the conservatives) had little or no awareness of it. Obviously, those not aware of the dominant community trend could not be using the community as a reference group for an attitude. (It does not follow, of course, that all those who are aware of it are necessarily using the community as reference group.) A simple measure of awareness was therefore devised. Subjects were asked to respond in two ways to a number of attitude statements taken from the P.E.P. scale: first, to indicate agreement or disagreement (for example, with the statement: 'The budget should be balanced before the government spends any

money on social security'); and second, to estimate what percentage of freshmen, juniors and seniors, and faculty would agree with the statement. From these responses was computed an index of divergence (of own attitude) from the estimated majority of juniors and seniors. Thus a positive index on the part of a senior indicates the degree to which her own responses are more conservative than those of her classmates, and a negative index the degree to which they are less conservative. Those seniors whose divergence index more or less faithfully reflects the true difference between own and class attitude may (or may not) be using the class as an attitude reference group; those whose divergence indexes represent an exaggerated or minimised version of the true relationship between own and class attitude are clearly not using the class as an attitude reference group, or if so, only in a fictitious sense. (For present purposes the junior–senior group may be taken as representative of the entire student body, since it is the group which 'sets the tone' of the total membership group.)

These data were supplemented by direct information obtained in interviews with seniors in three consecutive classes, just prior to graduation. Questions were asked about resemblance between own attitudes and those of class majorities and leaders, about parents' attitudes and own resemblance to them, about any alleged 'social pressure to become liberal', about probable reaction if the dominant college influence had been conservative instead of liberal, etc. Abundant information was also available from the college personnel office and from the college psychiatrist. It was not possible to combine all of these sources of information into intensive studies of each individual, but complete data were assembled for (roughly) the most conservative and least conservative sixths of three consecutive graduating classes. The twenty-four non-conservative and nineteen conservative seniors thus selected for intensive study were classified according to their indexes of conservative divergence and of community reputation. Thus eight sets of seniors were identified, all individuals within each set having in common similar attitude scores, similar reputations for community identification, and similar degrees of awareness (based upon divergence index) of own attitude position relative to classmates. The following descriptions of these eight sets of seniors will show that there was a characteristic pattern of relationship between membership group and reference group within each of the sets.

1. *Conservatives, reputedly negativistic, aware of their own relative conservatism.* Four of the five are considered stubborn or resistant by teachers (all five, by student judges). Three have prestige scores of 0, scores of the other two being about average for their class. Four of the five are considered by teachers or psychiatrist, or by both, to be overdependent upon one or both parents. All of the four who were interviewed described *their major hopes*, on entering college, *in terms of social rather than academic prestige*; all four felt that they had been defeated in this aim. The following verbatim quotations are illustrative:

E2

Probably the feeling that (my instructors) didn't accept me led me to reject their opinions. (She estimates classmates as being only moderately less conservative than herself, but faculty as much less so.)

G32

I wouldn't care to be intimate with those so-called 'liberal' student leaders. (*She claims to be satisfied with a small group of friends.* She is chosen as friend, in a sociometric questionnaire responded to by all students, only twice, and reciprocates both choices; both are conservative students.)

F22

I wanted to disagree with all the noisy liberals, but I was afraid and I couldn't. *So I built up a wall inside me against what they said. I found I couldn't compete, so I decided to stick to my father's ideas. For at least two years I've been insulated against all college influences.* (She is chosen but once as a friend, and does not reciprocate that choice.)

Q10

(who rather early concluded that she had no chance of social success in college). It hurt me at first, but now I don't give a damn. *The things I really care about are mostly outside the college.* I think radicalism symbolises the college for me more than anything else. (Needless to say, she has no use for radicals.)

For these four individuals (and probably for the fifth also) the community serves as reference group in a *negative* sense, and the

home-and-family group in a positive sense. Thus their conservatism is dually reinforced.

2. *Conservatives, reputedly negativistic, unaware of their own relative conservatism.* All five are described by teachers, as well as by guess-who judges, to be stubborn or resistant. Four have prestige scores of 0, and the fifth a less than average score. Each reciprocated just one friendship choice. Four are considered insecure in social relationships, and all five are regarded as extremely dependent upon parents. In interviews four describe with considerable intensity, and the fifth with more moderation, pre-college experiences of rebuff, ostracism or isolation, and all describe their hopes, on entering college, in terms of making friends or avoiding rebuff rather than in terms of seeking prestige. All five felt that their (rather modest) aims had met with good success. Each of the five denies building up any resistance to the acceptance of liberal opinions (but two add that they would have resented any such pressure, if felt). Three believe that only small, special groups in the college have such opinions, while the other two describe themselves as just going their own way, *paying no attention to anything but their own little circles and their college work.* Typical quotations follow:

Q47

I'm a perfect middle-of-the-roader, neither enthusiast nor critic. I'd accept anything if they just let me alone. . . . I've made all the friends I want. (Only one of her friendship choices is reciprocated.)

Q19

In high school I was always thought of as my parents' daughter. I never felt really accepted for myself. . . . I wanted to make my own way here, socially, but independence from my family has never asserted itself in other ways. (According to guess-who ratings, she is highly resistant to faculty authority.)

L12

What I most wanted was to get over being a scared bunny. . . . I always resent doing the respectable thing just because it's the thing to do, but I didn't realise I was so different, politically. from my classmates. At least I agree with the few people I ever talk to about such matters. (Sociometric responses place her in a small, conservative group.)

Q81

I hated practically all my school life before coming here. I had the perfect inferiority complex, and I pulled out of school social life – out of fear. I didn't intend to repeat that mistake here. . . . I've just begun to be successful in winning friendships, and I've been blissfully happy here. (She is described by teachers as 'pathologically belligerent'; she receives more than the average number of friendship choices, but reciprocates only one of them.)

For these five individuals, who are negativistic in the sense of being near-isolates rather than rebels, the community does not serve as reference group for public attitudes. To some extent, their small friendship groups serve in this capacity, but in the main they still refer such areas of their lives to the home-and-family group. They are too absorbed in their own pursuits to use the total membership group as a reference group for most other purposes, too.

3. *Conservatives, not reputedly negativistic, aware of their own relative conservatism.* Three of the five are described by teachers as 'co-operative' and 'eager', and none as stubborn or resistant. Four are above average in prestige. Four are considered by teachers or by guess-who raters, or both, to retain very close parental ties. All four who were interviewed had more or less definite ambitions for leadership on coming to college, and all felt that they had been relatively successful – though, in the words of one of them, none ever attained the 'really top-notch positions'. All four are aware of conflict between parents and college community in respect to public attitudes, and all quite consciously decided to 'string along' with parents, feeling self-confident of holding their own in college in spite of being atypical in this respect. Sample quotations follow:

Q73

I'm all my mother has in the world. It's considered intellectually superior here to be liberal or radical. This puts me on the defensive, as I refuse to consider my mother beneath me intellectually, as so many other students do. Apart from this, I have loved every aspect of college life. (A popular girl, many of whose friends are among the non-conservative college leaders.)

Q78

I've come to realise how much my mother's happiness depends on me, and the best way I can help her is to do things with her at home as often as I can. This has resulted in my not getting the feel of the college in certain ways, and I know my general conservatism is one of those ways. But it has not been important enough to me to make me feel particularly left out. If you're genuine and inoffensive about your opinions, no one really minds here if you remain conservative. (Another popular girl, whose friends were found among many groups.)

F32

Family against faculty has been my struggle here. As soon as I felt really secure here I decided not to let the college atmosphere affect me too much. Every time I've tried to rebel against my family I've found out how terribly wrong I am, and so I've naturally kept to my parents' attitudes. (While not particularly popular, she shows no bitterness and considerable satisfaction over her college experience.)

Q35

I've been aware of a protective shell against radical ideas. When I found several of my best friends getting that way, I either had to go along or just shut out that area entirely. I couldn't respect myself if I had changed my opinions just for that reason, and so I almost deliberately lost interest – really, *it was out of fear of losing my friends.* (A very popular girl, with no trace of bitterness, who is not considered too dependent upon parents.)

For these five the total membership group does not serve as reference group in respect to public attitudes, but does so serve for most other purposes. At some stage in their college careers the conflict between college community and home and family as reference group for public attitudes was resolved in favour of the latter.

4. *Conservatives, not reputedly negativistic, not aware of their own relative conservatism.* All four are consistently described by teachers as conscientious and co-operative; three are considered over-docile and uncritical of authority. All are characterised by feelings of inferiority. All are low in prestige, two receiving scores of 0; all are low in friendship choices, but reciprocate most of these few choices. Two are described as in conflict about parental authority, and two as dependent and contented. All four recall considerable anxiety as to whether they would fit into the college community; all feel that they have succeeded better than they had expected. Sample statements from interviews follow:

D22

I'd like to think like the college leaders, but I'm not bold enough and I don't know enough. So the college trend means little to me; I didn't even realise how much more conservative I am than the others. *I guess my family influence has been strong enough to counterbalance the college influence.* (This girl was given to severe emotional upsets, and according to personnel records, felt 'alone and helpless except when with her parents'.)

M12

It isn't that I've been resisting any pressure to become liberal. The influences here didn't matter enough to resist, I guess. *All that's really important that has happened to me occurred outside of college,* and so I never became very susceptible to college influences. (*Following her engagement to be married, in her second year, she had 'practically retired' from community life.*)

Q68

If I'd had more time here I'd probably have caught on to the liberal drift here. But I've been horribly busy making money and trying to keep my college work up. *Politics and that sort of thing I've always associated with home instead of with the college.* (A 'town girl' of working-class parentage.)

Q70

Most juniors and seniors, if they really *get excited about their work, forget about such community enthusiasms as sending telegrams to Congressmen.* It was so important to me to be accepted, I mean intellectually, *that I naturally came to identify myself in every way with the group which gave me this sort of intellectual satisfaction.* (One of a small group of science majors, nearly all conservative, who professed no interests other than science and who were highly self-sufficient socially.)

For none of the four was the total membership group a reference group for public attitudes. Unlike the non-negativistic conservatives who are aware of their relative conservatism, they refer to the

total membership group for few if any other purposes. Like the negativistic conservatives who are unaware of their relative conservatism, their reference groups for public attitudes are almost exclusively those related to home and family.

5. *Non-conservatives, reputedly community-identified, aware of their relative non-conservatism.* Each of the seven is considered highly independent by teachers, particularly in intellectual activities; all but one are referred to as meticulous, perfectionist, or over-conscientious. Four are very high in prestige, two high, and one average; all are 'good group members', and all but one a 'leader'. None is considered over-dependent upon parents. All have come to an understanding with parents concerning their 'liberal' views; five have 'agreed to differ', and the other two describe one or both parents as 'very liberal'. All take their public attitudes seriously, in most cases expressing the feeling that they have bled and died to achieve them. Interview excerpts follow:

B72

I bend in the direction of community expectation – almost more than I want to. I constantly have to check myself to be sure it's real self-conviction and not just social respect. (An outstanding and deeply respected leader.)

M42

My family has always been liberal, but the influences here made me go further, and for a while I was pretty far left. Now I'm pretty much in agreement with my family again, but it's my own and it means a lot. It wouldn't be easy for me to have friends who are very conservative. (Her friendship choices are exclusively given to non-conservatives.)

E72

I had been allowed so much independence by my parents that I needed desperately to identify myself with an institution with which I could conform conscientiously. Bennington was perfect. I drank up everything the college had to offer, including social attitudes, though not uncritically. I've become active in radical groups and constructively critical of them. (Both during and after college she worked with C.I.O. unions.)

H32

I accepted liberal attitudes here because *I had always secretly felt that my family was narrow and intolerant, and because such attitudes had prestige value.* It was all part of my generally expanding personality – *I had never really been part of anything before.* I don't accept things without examining things, however, and I was sure I meant it before I changed. (One of those who has 'agreed to differ' with parents.)

Q43

It didn't take me long to see that liberal attitudes had prestige value. But all the time I felt inwardly superior to persons who want public acclaim. Once I had arrived at a feeling of personal security, I could see that it wasn't important – it wasn't enough. *So many people have no security at all. I became liberal at first because of its prestige value.* I remain so because the problems around which my liberalism centres are important. What I want now is to be effective in solving the problems. (Another conspicuous leader, active in and out of college in liberal movements.)

The total membership clearly serves as reference group for these individuals' changing attitudes, but by no means as the only one. For those whose parents are conservative, parents represent a negative reference group, from whom emancipation was gained via liberal attitudes. And for several of them the college community served as a bridge to outside liberal groups as points of reference.

6. *Non-conservatives, reputedly community-identified, not aware of their own relative non-conservatism.* The word *enthusiastic* appears constantly in the records of each of these six. All are considered eager, ambitious, hard-working and anxious to please. Four are very high in prestige, the other two about average. None is considered over-dependent upon parents, and only two are known to have suffered any particular conflict in achieving emancipation. Each one came to college with ambitions for leadership, and each professes extreme satisfaction with her college experience. Sample quotations follow:

Qx

Every influence I felt tended to push me in the liberal direction: my underdog complex, *my need to be independent of my parents and my anxiousness to be a leader here.*

Q61

I met a whole body of new information here; I took a deep breath and plunged. When I talked about it at home my family began to treat me as if I had an adult mind. *Then too, my new opinions gave me the reputation here of being open-minded and capable of change.* I think I could have got really radical but I found it wasn't the way to get prestige here. (She judges most of her classmates to be as non-conservative as herself.)

Q72

I take everything hard, and so of course I reacted hard to all the attitudes I found here. I'm 100 per cent enthusiastic about Bennington, and that includes liberalism (but not radicalism, though I used to think so). Now I know that you can't be an *extremist if you're really devoted to an institution*, whether it's a labour union or a college. (A conspicuous leader who, like most of the others in this set of six, *judges classmates to be only slightly more conservative than herself.*)

Q63

I came to college to get away from my family, who never had any respect for my mind. Becoming a radical meant thinking for myself and, figuratively, thumbing my nose at my family. *It also meant intellectual identification with the faculty and students that I most wanted to be like.* (She has always felt oppressed by parental respectability and sibling achievements.)

Q57

It's very simple. *I was so anxious to be accepted that I accepted the political complexion of the community here.* I just couldn't stand out against the crowd unless I had many friends and strong support. (Not a leader, but many close friends among leaders and non-conservatives.)

For these six, like the preceding seven, the membership group serves as reference group for public affairs. They differ from the preceding seven chiefly in that they are less sure of themselves and are careful 'not to go too far'. Hence they tend to repudiate 'radicalism', and to judge classmates as only slightly less conservative than themselves.

7. *Non-conservatives, not reputedly community-identified, aware of own relative non-conservatism.* Each of the six is described as highly independent and critical-minded. Four are consistently reported as intellectually outstanding, and the other two occasionally so. All describe their ambitions on coming to college in intellectual rather than in social terms. Four of the five who were interviewed stated that in a conservative college they would be 'even more radical than here'. Two are slightly above average in prestige, two below average, and two have zero scores. Three have gone through rather severe battles in the process of casting off what they regard as parental shackles; none is considered over-dependent upon parents. Sample interview excerpts follow:

Q7

All my life I've resented the protection of governesses and parents. What I most wanted here was the intellectual approval of teachers and the more advanced students. Then I found you can't be reactionary and be intellectually respectable. (Her traits of independence became more marked as she achieved academic distinction.)

Q21

I simply got filled with new ideas here, and the only possible formulation of all of them was to adopt a radical approach. *I can't see my own position in the world in any other terms. The easy superficiality with which so many prestige-hounds here get 'liberal' only forced me to think it out more intensely.* (A highly gifted girl, considered rather aloof.)

C32

I started rebelling against my pretty stuffy family before I came to college. I felt apart from freshmen here, because I was older. Then I caught on to faculty attempts to undermine prejudice. I took sides with the faculty immediately, against the immature freshmen. I crusaded about it. *It provided just what I needed by way of family rebellion*, and bolstered up my self-confidence, too. (A very bright girl, regarded as sharp-tongued and a bit haughty.)

J24

I'm easily influenced by people whom I respect, and the people who rescued me when I was down and out, intellectually, gave me a radical intellectual approach; they included both teachers and advanced students. *I'm not rebelling against anything.* I'm just doing

what I had to do to stand on my own feet intellectually. (Her academic work was poor as a freshman, but gradually became outstanding.)

For these six students it is not the total membership group, but dominant subgroups (faculty, advanced students) which at first served as positive reference groups, and for many of them the home groups served as a negative point of reference. Later, they developed extra-college reference groups (left-wing writers, etc.). In a secondary sense, however, the total membership group served as a negative point of reference – i.e. they regarded their non-conservatism as a mark of personal superiority.

8. *Non-conservatives, not reputedly community-identified, not aware of own relative non-conservatism.* Each of the five is considered hard-working, eager and enthusiastic but (especially during the first year or two) unsure of herself and too dependent upon instructors. They are 'good citizens', but in a distinctly retiring way. Two are above average in prestige, and the other three much below average. None of the five is considered over-dependent upon parents; two are known to have experienced a good deal of conflict in emancipating themselves. All regard themselves as 'pretty average persons', with a strong desire to conform; they describe their ambitions in terms of social acceptance instead of social or intellectual prestige. Sample excerpts follow:

E22

Social security is the focus of it all with me. I became steadily less conservative as long as I was *needing to gain in personal security, both with students and with faculty.* I developed some resentment against a few extreme radicals who don't really represent the college viewpoint, and that's why I changed my attitudes so far and no further. (A girl with a small personal following, otherwise not especially popular.)

D52

Of course there's social pressure here to give up your conservatism. I'm glad of it, because for me this became the *vehicle for achieving independence from my family.* So changing my attitudes has gone hand in hand with two *very important things: establishing my own independence and at the same time becoming a part of the college organism.* (She attributes the fact that her social attitudes changed, while those of her younger sister, also at the college, did not, to the fact that she had greater need both of family independence and of group support.)

Q6

I was ripe for developing liberal or even radical opinions because so many of my friends at home were doing the same thing. So it was really wonderful that I could agree with all the people I respected here and the same time move in the direction that my home friends were going. (A girl characterised by considerable personal instability at first, but showing marked improvement.)

Qy

I think my change of opinions has given me *intellectual and social self-respect at the same time.* I used to be too timid for words, and I never had an idea of my own. As I gradually became more successful in my work and made more friends, I came to feel that it didn't matter so much whether I agreed with my parents. It's all part of the feeling that I really belong here. (Much other evidence confirms this; she was lonely and pathetic at first, but really belonged later.)

These five provide the example *par excellence* of individuals who came to identify themselves with 'the community' and whose attitudes change *pari passu* with the growing sense of identity. Home-and-family groups served as supplementary points of reference, either positive or negative. To varying degrees, subgroups within the community served as focal points of reference. But, because of *their need to be accepted, it was primarily the membership group as such which served as reference group for these five.*

SUMMARY

In this community, as presumably in most others, all individuals belong to the total membership group, but such membership is not necessarily a point of reference for every form of social adaptation, e.g. for acquiring attitudes towards public issues. *Such attitudes, however, are not acquired in a social vacuum. Their acquisition is a function of relating oneself to some group or groups, positively or negatively.* In many cases (perhaps in all) the referring of social attitudes to one group negatively leads to referring them to another group

positively, or vice versa, so that the attitudes are dually reinforced.

An individual is, of course, 'typical' in respect to attitudes if the total membership group serves as a positive point of reference for that purpose, but 'typicality' may also result from the use of other reference groups. It does not follow from the fact that an individual is 'atypical' that the membership group does not serve for reference purposes; it may serve as negative reference group. Even if the membership group does not serve as reference group at all (as in the case of conservatives in this community who are unaware of the general freshman-to-senior trend), it cannot be concluded that attitude development is not a function of belonging to the total membership group. The unawareness of such individuals is itself a resultant adaptation of particular individuals to a particular membership group. The fact that such individuals continue to refer attitudes towards public issues primarily to home-and-family groups is, in part at least, a result of the kind of community in which they have membership.

In short, the Bennington findings seem to support the thesis that, in a community characterised by certain approved attitudes, the individual's attitude development is a function of the way in which he relates himself both to the total membership group and to one or more reference groups. . . .

SOME CONDITIONS OF OBEDIENCE AND DISOBEDIENCE TO AUTHORITY

Stanley Milgram

THE situation in which one agent commands another to hurt a third turns up time and again as a significant theme in human relations. It is powerfully expressed in the story of Abraham, who is commanded by God to kill his son. It is no accident that Kierkegaard, seeking to orient his thought to the central themes of human experience, chose Abraham's conflict as the springboard to his philosophy.

War too moves forward on the triad of an authority which commands a person to destroy the enemy, and perhaps all organised hostility may be viewed as a theme and variation on the three elements of authority, executant, and victim.[1] We describe an experimental programme, recently concluded at Yale University, in which a particular expression of this conflict is studied by experimental means.

In its most general form the problem may be defined thus: if X tells Y to hurt Z, under what conditions will Y carry out the command of X and under what conditions will he refuse? In the more limited form possible in laboratory research, the question becomes: if an experimenter tells a subject to hurt another person, under what conditions will the subject go along with this instruction, and under what conditions will he refuse to obey? The laboratory problem is not so much a dilution of the general statement as one concrete expression of the many particular forms this question may assume.

One aim of the research was to study behaviour in a strong situation of deep consequence to the participants, for the psychological forces operative in powerful and lifelike forms of the conflict may not be brought into play under diluted conditions.

This approach meant, first, that we had a special obligation to protect the welfare and dignity of the persons who took part in the study; subjects were, of necessity, placed in a difficult predicament, and steps had to be taken to ensure their well-being before they were discharged from the laboratory. Towards this end, a careful, post-experimental treatment was devised and has been carried through for subjects in all conditions.[2]

TERMINOLOGY

If Y follows the command of X we shall say that he has obeyed X; if he fails to carry out the command of X, we shall say that he has disobeyed X. The terms *to obey* and *to disobey*, as used here, refer to the subject's overt action only, and carry no implication for the motive or experiential states accompanying the action.[3]

To be sure, the everyday use of the word *obedience* is not entirely free from complexities. It refers to action within widely varying situations, and connotes diverse motives within those situations: a child's obedience differs from a soldier's obedience, or the love, honour and *obey* of the marriage vow. However, a consistent behaviour relationship is indicated in most uses of the term: in the act of obeying, a person does what another person tells him to do.

Y obeys X if he carries out the prescription for action which X has addressed to him; the term suggests, moreover, that some form of dominance–subordination, or hierarchical element, is part of the situation in which the transaction between X and Y occurs.

A subject who complies with the entire series of experimental commands will be termed an *obedient* subject; one who at any point in the command series defies the experimenter will be called a *disobedient* or *defiant* subject. As used in this report, the terms refer only to the subject's performance in the experiment, and do not necessarily imply a general personality disposition to submit to or reject authority.

SUBJECT POPULATION

The subjects used in all experimental conditions were male adults, residing in the greater New Haven and Bridgeport areas, aged twenty to fifty years and engaged in a wide variety of occupations. Each experimental condition described in this report employed 40 fresh subjects and was carefully balanced for age and occupational types. The occupational composition for each experiment was: workers, skilled and unskilled: 40 per cent; white collar, sales, business: 40 per cent; professionals: 20 per cent. The occupations were intersected with three age categories (subjects in twenties, thirties and forties, assigned to each condition in the proportions of 20, 40 and 40 per cent respectively).

THE GENERAL LABORATORY PROCEDURE[4]

The focus of the study concerns the amount of electric shock a subject is willing to administer to another person when ordered by an experimenter to give the 'victim' increasingly more severe punishment. The act of administering shock is set in the context of a learning experiment, ostensibly designed to study the effect of punishment on memory. Aside from the experimenter, one naïve subject and one accomplice perform in each session. On arrival each subject is paid $4.50. After a general talk by the experimenter, telling how little scientists know about the effect of punishment on memory, subjects are informed that one member of the pair will serve as teacher and one as learner. A rigged drawing is held so that the naïve subject is always the teacher, and the accomplice becomes the learner. The learner is taken to an adjacent room and strapped into an 'electric chair'.

The naïve subject is told that it is his task to teach the learner a list of paired associates, to test him on the list, and to administer punishment whenever the learner errs in the test. Punishment takes the form of electric shock, delivered to the learner by means of a shock generator controlled by the naïve subject. The teacher is instructed to increase the intensity of electric shock one step on the generator on each error. The learner, according to plan, provides many wrong answers, so that before long the naïve subject must give him the strongest shock on the generator. Increases in shock level are met by increasingly insistent demands from the learner that the experiment be stopped because of the growing discomfort to him. However, in clear terms the experimenter orders the teacher to continue with the procedure in disregard of the learner's protests. *Thus, the naïve subject must resolve a conflict between two mutually incompatible demands from the social field. He may continue to follow the orders of the experimenter and shock the learner with increasing severity, or he may refuse to follow the orders of the experimenter and heed the learner's pleas. The experimenter's authority operates not in a free field, but against ever-mounting countervailing pressures from the person being punished.*

A few technical points need to be considered before the findings of the study are described. For the purpose of delivering shock, a simulated shock generator is used, with 30 clearly marked voltage levels that range from 15 to 450 volts (and each activated by one of 30 individual lever switches set in a horizontal line). The naïve subject is given a sample shock of 45 volts to convince him of the authenticity of the instrument. The generator bears verbal designations that range from 'Slight Shock' to 'Danger: Severe Shock'.

The responses of the victim are standardised on tape, and each protest is co-ordinated to a particular voltage level on the shock generator. Starting with 75 volts the learner begins to grunt and moan. At 150 volts he demands to be let out of the experiment. At 180 volts he cries out that he can no longer stand the pain. At 300 volts he refuses to provide any more answers to the memory test, insisting that he is no longer a participant in the experiment and must be freed. In response to this last tactic, the experimenter instructs the naïve subject to treat the absence of an answer as

equivalent to a wrong answer, and to follow the usual shock procedure. The experimenter reinforces his demand with the statement: 'You have no other choice, you must go on!' (This imperative is used whenever the naïve subject tries to break off the experiment.) If the subject refuses to give the next higher level of shock, the experiment is considered at an end. A quantitative value is assigned to the subject's performance based on the maximum intensity shock he administered before breaking off. Thus any subject's score may range from zero (for a subject unwilling to administer the first shock level) to 30 (for a subject who proceeds to the highest voltage level on the board). For any particular subject and for any particular experimental condition the degree to which participants have followed the experimenter's orders may be specified with a numerical value, corresponding to the metric on the shock generator.

This laboratory situation gives us a framework in which to study the subject's reactions to the principal conflict of the experiment. Again, this conflict is between the experimenter's demands that he continue to administer the electric shock, and the learner's demands, which become increasingly more insistent, that the experiment be stopped. The crux of the study is to vary systematically the factors believed to alter the degree of obedience to the experimental commands, to learn under what conditions submission to authority is most probable, and under what conditions defiance is brought to the fore.

PILOT STUDIES

Pilot studies for the present research were completed in the winter of 1960; they differed from the regular experiments in a few details; for one, the victim was placed behind a silvered glass, with the light balance on the glass such that the victim could be dimly perceived by the subject.

Though essentially qualitative in treatment, these studies pointed to several significant features of the experimental situation. At first no vocal feedback was used from the victim. It was thought that the verbal and voltage designations on the control panel would create sufficient pressure to curtail the subject's obedience. However, this was not the case. In the absence of protests from the learner, virtually all subjects, once commanded, went blithely to the end of the board, seemingly indifferent to the verbal designations ('Extreme Shock' and 'Danger: Severe Shock'). This deprived us of an adequate basis for scaling obedient tendencies. A force had to be introduced that would strengthen the subject's resistance to the experimenter's commands, and reveal individual differences in terms of a distribution of break-off points.

This force took the form of protests from the victim. Initially, mild protests were used, but proved inadequate. Subsequently, more vehement protests were inserted into the experimental procedure. To our consternation, even the strongest protests from the victim did not prevent all subjects from administering the harshest punishment ordered by the experimenter; but the protests did lower the mean maximum shock somewhat and created some spread in the subject's performance; therefore, the victim's cries were standardised on tape and incorporated into the regular experimental procedure.

The situation did more than highlight the technical difficulties of finding a workable experimental procedure: it indicated that subjects would obey authority to a greater extent than we had supposed. It also pointed to the importance of feedback from the victim in controlling the subject's behaviour.

One further aspect of the pilot study was that subjects frequently averted their eyes from the person they were shocking, often turning their heads in an awkward and conspicuous manner. One subject explained: 'I didn't want to see the consequences of what I had done.' Observers wrote:

> . . . subjects showed a reluctance to look at the victim, whom they could see through the glass in front of them. When this fact was brought to their attention they indicated that it caused them discomfort to see the victim in agony. We note, however, that although the subject refuses to look at the victim, he continues to administer shocks.

This suggested that the salience of the victim may have, in some degree, regulated the subject's performance. If, in obeying the experimenter, the subject found it necessary to avoid scrutiny of the victim, would the converse be true? If the victim were rendered increasingly more salient to the subject, would obedience diminish? The first set of regular experiments was designed to answer this question.

IMMEDIACY OF THE VICTIM

This series consisted of four experimental conditions. In each condition the victim was brought 'psychologically' closer to the subject giving him shocks.

In the first condition (Remote Feedback) the victim was placed in another room and could not be heard or seen by the subject, except that, at 300 volts, he pounded on the wall in protest. After 315 volts he no longer answered or was heard from.

The second condition (Voice Feedback) was identical to the first except that voice protests were introduced. As in the first condition the victim was placed in an adjacent room, but his complaints could be heard clearly through a door left slightly ajar, and through the walls of the laboratory.[5]

The third experimental condition (Proximity) was similar to the second, except that the victim was now placed in the same room as the subject, and 1½ ft from him. Thus he was visible as well as audible, and voice cues were provided.

The fourth, and final, condition of this series (Touch-Proximity) was identical to the third, with this exception: the victim received a shock only when his hand rested on a shockplate. At the 150-volt level the victim again demanded to be let free and, in this condition, refused to place his hand on the shockplate. The experimenter ordered the naïve subject to force the victim's hand on to the plate. Thus obedience in this condition required that the subject have physical contact with the victim in order to give him punishment beyond the 150-volt level.

Forty adult subjects were studied in each condition. The data revealed that obedience was significantly reduced as the victim was rendered more immediate to the subject. The mean maximum shock for the conditions is shown in Fig. 31.1.

Expressed in terms of the proportion of obedient to defiant subjects, the findings are that 34 per cent of the subjects defied the experimenter in the Remote condition, 37·5 per cent in Voice Feedback, 60 per cent in Proximity and 70 per cent in Touch-Proximity.

How are we to account for this effect? A first conjecture might be that as the victim was brought closer the subject became more aware of the intensity of his suffering and regulated his behaviour accordingly. This makes sense, but our evidence does not support

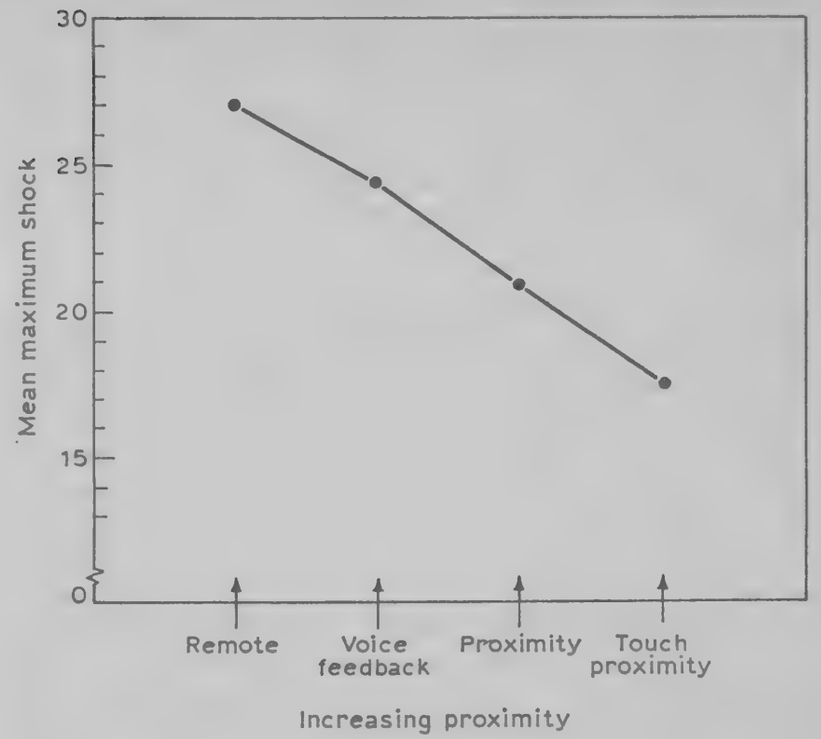


Fig. 31.1 Mean maxima in proximity series

the interpretation. There are no consistent differences in the attributed level of pain across the four conditions (i.e. the amount of pain experienced by the victim as estimated by the subject and expressed on a 14-point scale). But it is easy to speculate about alternative mechanisms:

Empathic cues

In the Remote and to a lesser extent the Voice Feedback condition, the victim's suffering possesses an abstract, remote quality for the subject. He is aware, but only in a conceptual sense, that his actions cause pain to another person; the fact is apprehended, but not felt. The phenomenon is common enough. The bombardier can reasonably suppose that his weapons will inflict suffering and death, yet this knowledge is divested of affect, and does not move him to a felt, emotional response to the suffering resulting from his actions. Similar observations have been made in wartime. It is

possible that the visual cues associated with the victim's suffering trigger empathic responses in the subject and provide him with a more complete grasp of the victim's experience. Or it is possible that the empathic responses are themselves unpleasant, possessing drive properties which cause the subject to terminate the arousal situation. Diminishing obedience, then, would be explained by the enrichment of empathic cues in the successive experimental conditions.

Denial and narrowing of the cognitive field

The Remote condition allows a narrowing of the cognitive field so that the victim is put out of mind. The subject no longer considers the act of depressing a lever relevant to moral judgement, for it is no longer associated with the victim's suffering. When the victim is close it is more difficult to exclude him phenomenologically. He necessarily intrudes on the subject's awareness since he is continuously visible. In the Remote conditions his existence and reactions are made known only after the shock has been administered. The auditory feedback is sporadic and discontinuous. In the Proximity conditions his inclusion in the immediate visual field renders him a continuously salient element for the subject. The mechanism of denial can no longer be brought into play. One subject in the Remote condition said: 'It's funny how you really begin to forget that there's a guy out there, even though you can hear him. For a long time I just concentrated on pressing the switches and reading the words.'

Reciprocal fields

If in the Proximity condition the subject is in an improved position to observe the victim, the reverse is also true. The actions of the subject now come under proximal scrutiny by the victim. Possibly, it is easier to harm a person when he is unable to observe our actions than when he can see what we are doing. His surveillance of the action directed against him may give rise to shame, or guilt, which may then serve to curtail the action. Many expressions of language refer to the discomfort or inhibitions that arise in face-to-face confrontation. It is often said that it is easier to criticise a man 'behind his back' than to 'attack him to his face'. If we are in the process of lying to a person it is reputedly difficult to 'stare him in the eye'. We 'turn away from others in shame' or in 'embarrassment' and this action serves to reduce our discomfort. The manifest function of allowing the victim of a firing squad to be blindfolded is to make the occasion less stressful for him, but it may also serve a latent function of reducing the stress of the executioner. In short, in the Proximity conditions, the subject may sense that he has become more salient in the victim's field of awareness. Possibly he becomes more self-conscious, embarrassed, and inhibited in his punishment of the victim.

Phenomenal unity of act

In the Remote conditions it is more difficult for the subject to gain a sense of *relatedness* between his own actions and the consequences of these actions for the victim. There is a physical and spatial separation of the act and its consequences. The subject depresses a lever in one room, and protests and cries are heard from another. The two events are in correlation, yet they lack a compelling phenomenological unity. The structure of a meaningful act – *I am hurting a man* – breaks down because of the spatial arrangements, in a manner somewhat analogous to the disappearance of phi phenomena when the blinking lights are spaced too far apart. The unity is more fully achieved in the Proximity conditions as the victim is brought closer to the action that causes him pain. It is rendered complete in Touch-Proximity.

Incipient group formation

Placing the victim in another room not only takes him further from the subject, but the subject and the experimenter are drawn relatively closer. There is incipient group formation between the experimenter and the subject, from which the victim is excluded. The wall between the victim and the others deprives him of an intimacy which the experimenter and subject feel. In the Remote condition, the victim is truly an outsider, who stands alone, physically and psychologically.

When the victim is placed close to the subject, it becomes easier to form an alliance with him against the experimenter. Subjects no longer have to face the experimenter alone. They have an ally who is close at hand and eager to collaborate in a revolt against the experimenter. Thus, the changing set of spatial relations leads to a potentially shifting set of alliances over the several experimental conditions.

Acquired behaviour dispositions

It is commonly observed that laboratory mice will rarely fight with their litter mates. Scott explains this in terms of passive inhibition. He writes: 'By doing nothing under . . . circumstances [the animal] learns to do nothing, and this may be spoken of as passive inhibition . . . this principle has great importance in teaching an individual to be peaceful, for it means that he can learn not to fight simply by not fighting.' Similarly, we may learn not to harm others simply by not harming them in everyday life. Yet this learning occurs in a context of proximal relations with others, and may not be generalised to that situation in which the person is physically removed from us. Or possibly, in the past, aggressive actions against others who were physically close resulted in retaliatory punishment which extinguished the original form of response. In contrast, aggression against others at a distance may have only sporadically led to retaliation. Thus the organism learns that it is safer to be aggressive towards others at a distance, and precarious to be so when the parties are within arm's reach. Through a pattern of rewards and punishments, he acquires a disposition to avoid aggression at close quarters, a disposition which does not extend to harming others at a distance. And this may account for experimental findings in the remote and proximal experiments.

Proximity as a variable in psychological research has received far less attention than it deserves. If men were sessile it would be easy to understand this neglect. But we move about; our spatial relations shift from one situation to the next, and the fact that we are near or remote may have a powerful effect on the psychological processes that mediate our behaviour towards others. In the present situation, as the victim is brought closer to the man ordered to give him shocks, increasing numbers of subjects break off the experiment, refusing to obey. The concrete, visible and proximal presence of the victim acts in an important way to counteract the experimenter's power and to generate disobedience.[6]

CLOSENESS OF AUTHORITY

If the spatial relationship of the subject and victim is relevant to the degree of obedience, would not the relationship of subject to experimenter also play a part?

There are reasons to feel that, on arrival, the subject is oriented primarily to the experimenter rather than to the victim. He has come to the laboratory to fit into the structure that the experimenter – not the victim – would provide. He has come less to understand his behaviour than to *reveal* that behaviour to a competent scientist, and he is willing to display himself as the scientist's purposes require. Most subjects seem quite concerned about the appearance they are making before the experimenter, and one could argue that this preoccupation in a relatively new and strange setting makes the subject somewhat insensitive to the triadic nature of the social situation. In other words, the subject is so concerned about the show he is putting on for the experimenter that influences from other parts of the social field do not receive as much weight as they ordinarily would. This overdetermined orientation to the experimenter would account for the relative insensitivity of the subject to the victim, and would also lead us to believe that alterations in the relationship between subject and experimenter would have important consequences for obedience.

In a series of experiments we varied the physical closeness and degree of surveillance of the experimenter. In one condition the experimenter sat just a few feet away from the subject. In a second condition, after giving initial instructions, the experimenter left the laboratory and gave his orders by telephone; in still a third condition the experimenter was never seen, providing instructions by means of a tape recording activated when the subjects entered the laboratory.

Obedience dropped sharply as the experimenter was physically removed from the laboratory. The number of obedient subjects in the first condition (Experimenter Present) was almost three times as great as in the second, where the experimenter gave his orders by telephone. Twenty-six subjects were fully obedient in the first condition, and only 9 in the second (Chi square obedient *vs.* defiant in the two conditions, 1 d.f. $= 14{\cdot}7$; $p < {\cdot}001$). Subjects seemed able to take a far stronger stand against the experimenter when they did not have to encounter him face to face, and the experimenter's power over the subject was severely curtailed.[7]

Moreover, when the experimenter was absent, subjects displayed an interesting form of behaviour that had not occurred under his surveillance. Though continuing with the experiment, several subjects administered lower shocks than were required

and never informed the experimenter of their deviation from the correct procedure. (Unknown to the subjects, shock levels were automatically recorded by an Esterline–Angus event recorder wired directly into the shock generator; the instrument provided us with an objective record of the subject's performance.) Indeed, in telephone conversations some subjects specifically assured the experimenter that they were raising the shock level according to instruction, whereas in fact they were repeatedly using the lowest shock on the board. This form of behaviour is particularly interesting: although these subjects acted in a way that clearly undermined the avowed purposes of the experiment, they found it easier to handle the conflict in this manner than to precipitate an open break with authority.

Other conditions were completed in which the experimenter was absent during the first segment of the experiment, but reappeared at the point that the subject definitely refused to give higher shocks when commanded by telephone. Although he had exhausted his power via telephone, the experimenter could frequently force further obedience when he reappeared in the laboratory.

Experiments in this series show that the physical *presence* of an authority is an important force contributing to the subject's obedience or defiance. Taken together with the first experimental series on the proximity of the victim, it would appear that something akin to fields of force, diminishing in effectiveness with increasing psychological distance from their source, have a controlling effect on the subject's performance. As the victim is brought closer, the subject finds it harder to administer shocks to him. When the victim's position is held constant relative to the subject, and the authority is made more remote, the subject finds it easier to break off the experiment. This effect is substantial in both cases, but manipulation of the experimenter's position yielded the more powerful results. Obedience to destructive commands is highly dependent on the proximal relations between authority and subject.

TENSIONS

The description of final scores does not fully convey the character of the subjects' performance, and it would be useful to interrupt our reporting of quantitative relationships to remark on the subjects' general reaction to the situation.

One might suppose that a subject would simply break off or continue as his conscience and temperament dictated. Yet this is very far from what happened. There were powerful reactions of tension and emotional strain in a substantial proportion of the participants. Persons were observed to sweat, tremble, stutter, bite their lips and groan as they found themselves increasingly implicated in the experimental conflict. Some sense of the subjects' mood can be got from a transcript of their remarks at break-off point. Here are the remarks of two defiant subjects:

> Is he banging? . . . Is he hurt out there? Well, I don't want to be responsible for anything that happens to him. No, I can't go on with it. I don't know whether he's all right or not. I mean he may have a heart condition or something; I wouldn't feel right doing it. He's banging for help. He's evidently in trouble or he wouldn't bang . . . I don't see any sense to this – to knocking him out or something. I just can't see it.
> (*Subject refuses to administer shocks*) [*Subject 0108*]

> The man is banging. He wants to get out . . . I'm sorry. If the man wants to get out I'd just as soon stop . . . I don't want to administer any more [extremely excited] . . . I will gladly refund the cheque, but I will not harm this man . . . I'm sorry, I will not give him any more if he obviously wants to get out. There is no money in the world that will make me hurt another individual. [*Subject 010X*]

And here is a transcript from an obedient subject. He began the experiment calmly, but became increasingly tense as the experiment proceeded. After administering the 150-volt shock, he began to address the experimenter in agitated tones:

> *150 volts delivered.* You want me to keep going?
> *165 volts delivered.* That guy is hollering in there. There's a lot of them here. He's liable to have heart condition. You want me to go?
> *180 volts delivered.* He can't stand it! I'm not going to kill that man in there! You hear him hollering? He's hollering. He can't stand it. What if something happens to him? . . . I'm not going to get that man sick in there. He's hollering in there. You know

what I mean? I mean I refuse to take responsibility. He's getting hurt in there. He's in there hollering. Too many left here. Geez, if he gets them wrong. There's too many of them left. I mean who is going to take responsibility if anything happens to that gentleman?

[*The experimenter accepts responsibility.*] All right.

195 volts delivered. You see he's hollering. Hear that. Gee, I don't know. [*The experimenter says: 'The experiment requires that you go on'.*] – I know it does, sir, but I mean – hugh – he don't know what he's in for. He's up to 195 volts.

210 volts delivered.

225 volts delivered.

240 volts delivered. Aw, no. You mean I've got to keep going up with the scale? No sir. I'm not going to kill that man! I'm not going to give him 450 volts! [*The experimenter says: 'The experiment requires that you go on.'* – I know it does, but that man is hollering in there, sir . . .

Despite his numerous, agitated objections, which were constant accompaniments to his actions, the subject unfailingly obeyed the experimenter, proceeding to the highest shock level on the generator. He displayed a curious dissociation between word and action. Although at the verbal level he had resolved not to go on, his actions were fully in accord with the experimenter's commands. This subject did not want to shock the victim, and he found it an extremely disagreeable task, but he was unable to invent a response that would free him from *E*'s authority. Many subjects cannot find the specific verbal formula that would enable them to reject the role assigned to them by the experimenter. Perhaps our culture does not provide adequate models for disobedience.

One puzzling sign of tension was the regular occurrence of nervous laughing fits. In the first four conditions 71 of the 160 subjects showed definite signs of nervous laughter and smiling. The laughter seemed entirely out of place, even bizarre. Full-blown, uncontrollable seizures were observed for 15 of these subjects. On one occasion we observed a seizure so violently convulsive that it was necessary to call a halt to the experiment. In the post-experimental interviews subjects took pains to point out that they were not sadistic types and that the laughter did not mean they enjoyed shocking the victim.

In the interview following the experiment subjects were asked to indicate on a 14-point scale just how nervous or tense they felt at the point of maximum tension (Fig. 31.2). The scale ranged from 'Not at all tense and nervous' to 'Extremely tense and nervous'. Self-reports of this sort are of limited precision, and at

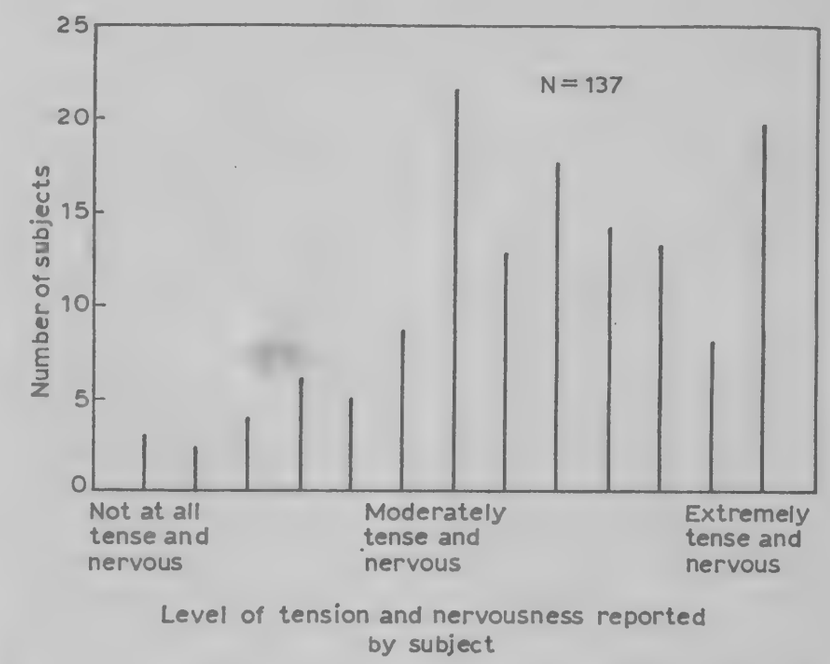


Fig. 31.2 Level of tension and nervousness, showing the self-reports on 'tension and nervousness' for 137 subjects in the proximity experiments. Subjects were given a scale with fourteen values ranging from 'Not at all tense and nervous' to 'Extremely tense and nervous'. They were instructed: 'Thinking back to that point in the experiment when you felt the most tense and nervous, indicate just how you felt by placing an X at the appropriate point of the scale.' The results are shown in terms of mid-point values.

best provide only a rough indication of the subject's emotional response. Still, taking the reports for what they are worth, it can be seen that the distribution of responses spans the entire range of the scale, with the majority of subjects concentrated at the centre and upper extreme. A further breakdown showed that obedient subjects reported themselves as having been slightly more tense

and nervous than the defiant subjects at the point of maximum tension.

How is the occurrence of tension to be interpreted? First, it points to the presence of conflict. If a tendency to comply with authority were the only psychological force operating in the situation, all subjects would have continued to the end and there would have been no tension. Tension, it is assumed, results from the simultaneous presence of two or more incompatible response tendencies. If sympathetic concern for the victim were the exclusive force, all subjects would have calmly defied the experimenter. Instead, there were both obedient and defiant outcomes, frequently accompanied by extreme tension. A conflict develops between the deeply ingrained disposition not to harm others and the equally compelling tendency to obey others who are in authority. The subject is quickly drawn into a dilemma of a deeply dynamic character, and the presence of high tension points to the considerable strength of each of the antagonistic vectors.

Moreover, tension defines the strength of the aversive state from which the subject is unable to escape through disobedience. When a person is uncomfortable, tense or stressed, he tries to take some action that will allow him to terminate this unpleasant state. Thus tension may serve as a drive that leads to escape behaviour. But in the present situation, even where tension is extreme, many subjects are unable to perform the response that will bring about relief. Therefore there must be a competing drive, tendency or inhibition that precludes activation of the disobedient response. The strength of this inhibiting factor must be of greater magnitude than the stress experienced, else the terminating act would occur. Every evidence of extreme tension is at the same time an indication of the strength of the forces that keep the subject in the situation.

Finally, tension may be taken as evidence of the reality of the situations for the subjects. Normal subjects do not tremble and sweat unless they are implicated in a deep and genuinely felt predicament.

BACKGROUND AUTHORITY

In psychophysics, animal learning, and other branches of psychology, the fact that measures are obtained at one institution rather than another is irrelevant to the interpretation of the findings, so long as the technical facilities for measurement are adequate and the operations are carried out with competence.

But it cannot be assumed that this holds true for the present study. The effectiveness of the experimenter's commands may depend in an important way on the larger institutional context in which they are issued. The experiments described thus far were conducted at Yale University, an organisation which most subjects regarded with respect and sometimes awe. In post-experimental interviews several participants remarked that the locale and sponsorship of the study gave them confidence in the integrity, competence, and benign purposes of the personnel; many indicated that they would not have shocked the learner if the experiments had been done elsewhere.

This issue of background authority seemed to us important for an interpretation of the results that had been obtained thus far; moreover it is highly relevant to any comprehensive theory of human obedience. Consider, for example, how closely our compliance with the imperatives of others is tied to particular institutions and locales in our day-to-day activities. On request, we expose our throats to a man with a razor blade in the barber shop, but would not do in a shoe store; in the latter setting we willingly follow the clerk's request to stand in our stockinged feet, but resist the command in a bank. In the laboratory of a great university, subjects may comply with a set of commands that would be resisted if given elsewhere. *One must always question the relationship to a person's sense of the context in which he is operating.*

To explore the problem we moved our apparatus to an office building in industrial Bridgeport and replicated experimental conditions, without any visible tie to the university.

Bridgeport subjects were invited to the experiment through a mail circular similar to the one used in the Yale study, with appropriate changes in letterhead, etc. As in the earlier study, subjects were paid $4.50 for coming to the laboratory. The same age and occupational distributions used at Yale, and the identical personnel, were employed.

The purpose in relocating in Bridgeport was to assure a complete dissociation from Yale, and in this regard we were fully successful. On the surface, the study appeared to be conducted by RESEARCH ASSOCIATES OF BRIDGEPORT, an organisation of unknown

character (the title had been concocted exclusively for use in this study).

The experiments were conducted in a three-room office suite in a somewhat run-down commercial building located in the downtown shopping area. The laboratory was sparsely furnished, though clean and marginally respectable in appearance. When subjects inquired about professional affiliations, they were informed only that we were a private firm conducting research for industry.

Some subjects displayed a scepticism concerning the motives of the Bridgeport experimenter. One gentleman gave us a written account of the thoughts he experienced at the control board

> . . . Should I quit this damn test? Maybe he passed out? What dopes we were not to check up on this deal. How do we know that these guys are legit? No furniture, bare walls, no telephone. We could of called the Police up or the Better Business Bureau. I learned a lesson tonight. How do I know that Mr Williams [the experimenter] is telling the truth . . . I wish I knew how many volts a person could take before lapsing into unconsciousness . . .
>
> [*Subject 2414*]

Another subject stated:

> I questioned on my arrival my own judgement [about coming]. I had doubts as to the legitimacy of the operation and the consequences of participation. I felt it was a heartless way to conduct memory or learning processes on human beings and certainly dangerous without the presence of a medical doctor.
>
> [*Subject 2440 V*]

There was no noticeable reduction in tension for the Bridgeport subjects. And the subjects' estimation of the amount of pain felt by the victim was slightly, though not significantly, higher than in the Yale study.

A failure to obtain complete obedience in Bridgeport would indicate that the extreme compliance found in New Haven subjects was tied closely to the background authority of Yale University; if a large proportion of the subjects remained fully obedient, very different conclusions would be called for.

As it turned out, the level of obedience in Bridgeport, although somewhat reduced, was not significantly lower than that obtained at Yale. A large proportion of the Bridgeport subjects were fully obedient to the experimenter's commands (48 per cent of the Bridgeport subjects delivered the maximum shock *vs.* 65 per cent in the corresponding condition at Yale).

How are these findings to be interpreted? It is possible that if commands of a potentially harmful or destructive sort are to be perceived as legitimate they must occur within some sort of institutional structure. But it is clear from the study that it need not be a particularly reputable or distinguished institution. The Bridgeport experiments were conducted by an unimpressive firm lacking any credentials; the laboratory was set up in a respectable office building with title listed in the building directory. Beyond that, there was no evidence of benevolence or competence. It is possible that the *category* of institution, judged according to its professed function, rather than its qualitative position within that category, wins our compliance. Persons deposit money in elegant, but also in seedy-looking banks, without giving much thought to the differences in security they offer. Similarly, our subjects may consider one laboratory to be as competent as another, so long as it *is* a scientific laboratory.

It would be valuable to study the subjects' performance in other contexts which go even further than the Bridgeport study in denying institutional support to the experimenter. It is possible that, beyond a certain point, obedience disappears completely. But that point had not been reached in the Bridgeport office: almost half the subjects obeyed the experimenter fully.

FURTHER EXPERIMENTS

We may mention briefly some additional experiments undertaken in the Yale series. A considerable amount of obedience and defiance in everyday life occurs in connection with groups. And we had reason to feel in the light of many group studies already done in psychology that group forces would have a profound effect on reactions to authority. A series of experiments was run to examine these effects. In all cases only one naïve subject was studied per hour, but he performed in the midst of actors who, unknown to him, were employed by the experimenter. In one experiment (Groups for Disobedience) two actors broke off in the middle of

the experiment. When this happened 90 per cent of the subjects followed suit and defied the experimenter. In another condition the actors followed the orders obediently; this strengthened the experimenter's power only slightly. In still a third experiment the job of pushing the switch to shock the learner was given to one of the actors, while the naïve subject performed a subsidiary act. We wanted to see how the teacher would respond if he were involved in the situation but did not actually give the shocks. In this situation only three subjects out of forty broke off. In a final group experiment the subjects themselves determined the shock level they were going to use. Two actors suggested higher and higher shock levels; some subjects insisted, despite group pressure, that the shock level be kept low; others followed along with the group.

Further experiments were completed using women as subjects, as well as a set dealing with the effects of dual, unsanctioned and conflicting authority. A final experiment concerned the personal relationship between victim and subject. These will have to be described elsewhere, lest the present report by extended to monographic length.

It goes without saying that future research can proceed in many different directions. What kinds of response from the victim are most effective in causing disobedience in the subject? Perhaps passive resistance is more effective than vehement protest. What conditions of entry into an authority system lead to greater or lesser obedience? What is the effect of anonymity and masking on the subject's behaviour? What conditions lead to the subject's perception of responsibility for his own actions? Each of these could be a major research topic in itself, and can readily be incorporated into the general experimental procedure described here.

LEVELS OF OBEDIENCE AND DEFIANCE

One general finding that merits attention is the high level of obedience manifested in the experimental situation. Subjects often expressed deep disapproval of shocking a man in the face of his objections, and others denounced it as senseless and stupid. Yet many subjects complied even while they protested. The proportion of obedient subjects greatly exceeded the expectations of the experimenter and his colleagues. At the outset, we had conjectured that subjects would not, in general, go above the level of 'Strong Shock'. In practice, many subjects were willing to administer the most extreme shocks available when commanded by the experimenter. For some subjects the experiment provides an occasion for aggressive release. And for others it demonstrates the extent to which obedient dispositions are deeply ingrained, and are engaged irrespective of their consequences for others. Yet this is not the whole story. Somehow, the subject becomes implicated in a situation from which he cannot disengage himself.

The departure of the experimental results from intelligent expectation, to some extent, has been formalised. The procedure was to describe the experimental situation in concrete detail to a group of competent persons, and to ask them to predict the performance of 100 hypothetical subjects. For purposes of indicating the distribution of break-off points judges were provided with a diagram of the shock generator, and recorded their predictions before being informed of the actual results. Judges typically underestimated the amount of obedience demonstrated by subjects.

In Fig. 31.3, we compare the predictions of forty psychiatrists at a leading medical school with the actual performance of subjects in the experiment. The psychiatrists predicted that most subjects would not go beyond the tenth shock level (150 volts; at this point the victim makes his first explicit demand to be freed). They further predicted that by the twentieth shock level (300 volts; the victim refuses to answer) 3·73 per cent of the subjects would still be obedient; and that only a little over one-tenth of one per cent of the subjects would administer the highest shock on the board. But, as the graph indicates, the obtained behaviour was very different. Sixty-two per cent of the subjects obeyed the experimenter's commands fully. Between expectation and occurrence there is a whopping discrepancy.

Why did the psychiatrists underestimate the level of obedience? Possibly, because their predictions were based on an inadequate conception of the determinants of human action, a conception that focuses on motives *in vacuo*. This orientation may be entirely adequate for the repair of bruised impulses as revealed on the psychiatrist's couch, but as soon as our interest turns to action in larger settings, attention must be paid to the situations in which motives are expressed. A situation exerts an important press on the individual. It exercises constraints and may provide push. In

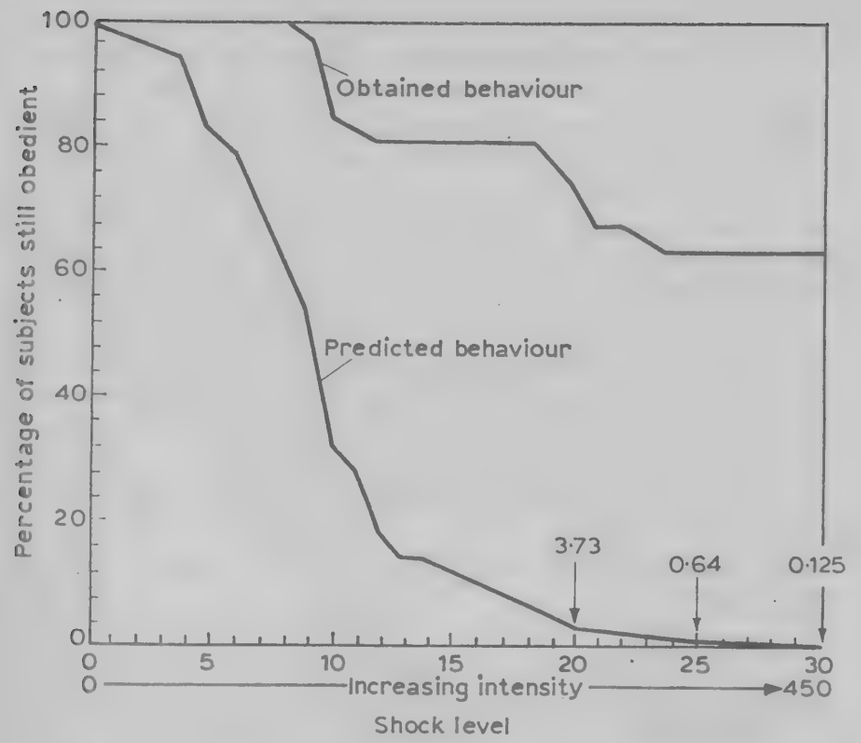


Fig. 31.3 Predicted and obtained behaviour in voice feedback

certain circumstances it is not so much the kind of person a man is, as the kind of situation in which he is placed, that determines his actions.

Many people, not knowing much about the experiment, claim that subjects who go to the end of the board are sadistic. Nothing could be more foolish as an overall characterisation of these persons. It is like saying that a person thrown into a swift-flowing stream is necessarily a fast swimmer, or that he has great stamina because he moves so rapidly relative to the bank. The context of action must always be considered. The individual, upon entering the laboratory, becomes integrated into a situation that carries its own momentum. The subject's problem then is how to become disengaged from a situation which is moving in an altogether ugly direction.

The fact that disengagement is so difficult testifies to the potency of the forces that keep the subject at the control board. Are these forces to be conceptualised as individual motives and expressed in the language of personality dynamics, or are they to be seen as the effects of social structure and pressures arising from the situation field?

A full understanding of the subject's action will, I feel, require that both perspectives be adopted. The person brings to the laboratory enduring dispositions towards authority and aggression, and at the same time he becomes enmeshed in a social structure that is no less an objective fact of the case. From the standpoint of personality theory one may ask: What mechanisms of personality enable a person to transfer responsibility to authority? What are the motives underlying obedient and disobedient performance? Does orientation to authority lead to a short-circuiting of the shame – guilt system? What cognitive and emotional defences are brought into play in the case of obedient and defiant subjects?

The present experiments are not, however, directed towards an exploration of the motives engaged when the subject obeys the experimenter's commands. Instead, they examine the situational variables responsible for the elicitation of obedience. Elsewhere, we have attempted to spell out some of the structural properties of the experimental situation that account for high obedience, and this analysis need not be repeated here. The experimental variations themselves represent our attempt to probe that structure, by systematically changing it and noting the consequences for behaviour. It is clear that some situations produce greater compliance with the experimenter's commands than others. However, this does not necessarily imply an increase or decrease in the strength of any single definable motive. Situations producing the greatest obedience could do so by triggering the most powerful, yet perhaps the most idiosyncratic, of motives in each subject confronted by the setting. Or they may simply recruit a greater number and variety of motives in their service. But whatever the motives involved – and it is far from certain that they can ever be known – action may be studied as a direct function of the situation in which it occurs. This has been the approach of the present study, where we sought to plot behavioural regularities against manipulated properties of the social field. Ultimately, social psychology would like to have a compelling *theory of situations* which will, first, present a language in terms of which situations can be defined; proceed to a typology of situations; and then point to the

manner in which definable properties of situations are transformed into psychological forces in the individual.

POSTSCRIPT

Almost a thousand adults were individually studied in the obedience research, and there were many specific conclusions regarding the variables that control obedience and disobedience to authority. Some of these have been discussed briefly in the preceding sections, and more detailed reports will be released subsequently.

There are now some other generalisations I should like to make, which do not derive in any strictly logical fashion from the experiments as carried out, but which, I feel, ought to be made. They are formulations of an intuitive sort that have been forced on me by observation of many subjects responding to the pressures of authority. The assertions represent a painful alteration in my thinking; and since they were acquired only under the repeated impact of direct observation, I have no illusion that they will be generally accepted by persons who have not had the same experience.

With numbing regularity good people were seen to knuckle under the demands of authority and perform actions that were callous and severe. Men who are in everyday life responsible and decent were seduced by the trappings of authority, by the control of their perceptions and by the uncritical acceptance of the experimenter's definition of the situation, into performing harsh acts.

What is the limit of such obedience? At many points we attempted to establish a boundary. Cries from the victim were inserted; not good enough. The victim claimed heart trouble; subjects still shocked him on command. The victim pleaded that he be let free, and his answers no longer registered on the signal box; subjects continued to shock him. At the outset we had not conceived that such drastic procedures would be needed to generate disobedience, and each step was added only as the ineffectiveness of the earlier techniques became clear. The final effort to establish a limit was the Touch–Proximity condition. But the very first subject in this condition subdued the victim on command, and proceeded to the highest shock level. A quarter of the subjects in this condition performed similarly.

The results, as seen and felt in the laboratory, are to this author disturbing. They raise the possibility that human nature, or – more specifically – the kind of character produced in American democratic society, cannot be counted on to insulate its citizens from brutality and inhumane treatment at the direction of malevolent authority. A substantial proportion of people do what they are told to do, irrespective of the content of the act and without limitations of conscience, so long as they perceive that the command comes from a legitimate authority. If in this study an anonymous experimenter could successfully command adults to subdue a fifty-year-old man, and force on him painful electric shocks against his protests, one can only wonder what government, with its vastly greater authority and prestige, can command of its subjects. There is, of course, the extremely important question of whether malevolent political institutions could or would arise in American society. The present research contributes nothing to this issue.

In an article entitled 'The Dangers of Obedience', Harold J. Laski wrote:[8]

> . . . civilisation means, above all, an unwillingness to inflict unnecessary pain. Within the ambit of that definition, those of us who heedlessly accept the commands of authority cannot yet claim to be civilised men.
>
> . . . Our business, if we desire to live a life, not utterly devoid of meaning and significance, is to accept nothing which contradicts our basic experience merely because it comes to us from tradition or convention or authority. It may well be that we shall be wrong; but our self-expression is thwarted at the root unless the certainties we are asked to accept coincide with the certainties we experience. That is why the condition of freedom in any state is always a widespread and consistent scepticism of the canons upon which power insists.

Notes

1. Consider, for example, J. P. Scott's analysis of war in his monograph on aggression:

'. . . while the actions of key individuals in a war may be explained in terms of direct stimulation to aggression, vast numbers of other people are involved simply by being part of an organised society.

'. . . For example, at the beginning of World War I an Austrian archduke was assassinated in Sarajevo. A few days later soldiers from all over Europe were marching towards each other, not because they were

stimulated by the archduke's misfortune, but because they had been trained to obey orders.' See J. P. Scott, *Aggression* (Chicago, 1958).

2. It consisted of an extended discussion with the experimenter and, of equal importance, a friendly reconciliation with the victim. It is made clear that the victim did not receive painful electric shocks. After the completion of the experimental series, subjects were sent a detailed report of the results and full purposes of the experimental programme. A formal assessment of this procedure points to its overall effectiveness. Of the subjects, 83·7 per cent indicated that they were glad to have taken part in the study; 15·1 per cent reported neutral feelings; and 1·3 per cent stated that they were sorry to have participated. A large number of subjects spontaneously requested that they be used in further experimentation. Four-fifths of the subjects felt that more experiments of this sort should be carried out, and 74 per cent indicated that they had learned something of personal importance as a result of being in the study. Furthermore, a university psychiatrist, experienced in outpatient treatment, interviewed a sample of experimental subjects with the aim of uncovering possible injurious effects resulting from participation. No such effects were in evidence. Indeed, subjects typically felt that their participation was instructive and enriching. S. Milgram, 'Issues in the Study of Obedience', *Amer. Psychol.*, vol. 19 (1964) 848–52.

3. *To obey* and *to disobey* are not the only terms one could use in describing the critical action of Y. One could say that Y is co-operating with X, or displays conformity with regard to X's commands. However, *co-operation* suggests that X agrees with Y's ends, and understands the relationship between his own behaviour and the attainment of those ends. (But the experimental procedure, and, in particular, the experimenter's command that the subject shock the victim even in the absence of a response from the victim, preclude such understanding.) Moreover, co-operation implies status parity for the co-acting agents, and neglects the asymmetrical, dominance–subordination element prominent in the laboratory relationship between experimenter and subject. *Conformity* has been used in other important contexts in social psychology, and most frequently refers to imitating the judgements or actions of others when no explicit requirement for imitation has been made. Furthermore, in the present study there are two sources of social pressure: pressure from the experimenter issuing the commands, and pressure from the victim to stop the punishment. It is the pitting of a common man (the victim) against an authority (the experimenter) that is the distinctive feature of the conflict. At a point in the experiment the victim demands that he be let free. The experimenter insists that the subject continue to administer shocks. Which act of the subject can be interpreted as conformity? The subject may conform to the wishes of his peer or to the wishes of the experimenter, and conformity in one direction means the absence of conformity in the other. Thus the word has no useful reference in this setting, for the dual and conflicting social pressures cancel out its meaning.

In the final analysis, the linguistic symbol representing the subject's action must take its meaning from the concrete context in which that action occurs; and there is probably no word in everyday language that covers the experimental situation exactly, without omissions or irrelevant connotations. It is partly for convenience, therefore, that the terms *obey* and *disobey* are used to describe the subject's actions. At the same time, our use of the words is highly congruent with dictionary meaning.

4. For more on laboratory procedures see S. Milgram, 'Behavioral Study of Obedience', *J. Abnorm. Soc. Psychol.*, vol. 67 (1963), 371–8. A similar and independently evolved experimental procedure, employing a shock generator, victim, and learning task, was reported by Buss. Buss used the technique for studying aggression, not obedience, and did not make use of the fundamental measure employed in the present study: break-off points. His investigation promises to be a useful complement to the present research. See A. H. Buss, *The Psychology of Aggression* (New York, 1961).

5. It is difficult to convey on the printed page the full tenor of the victim's responses, for we have no adequate notation for vocal intensity, timing and general qualities of delivery. Yet these features are crucial to producing the effect of an increasingly severe reaction to mounting voltage levels. (They can be communicated fully only by sending interested parties the recorded tapes.) In general terms, however, the victim indicates no discomfort until the 75-volt shock is administered, at which time there is a light grunt in response to the punishment. Similar reactions follow the 90- and 105-volt shocks, and at 120 volts the victim shouts to the experimenter that the shocks are becoming painful. Painful groans are heard on administration of the 135-volt shock, and at 150 volts the victim cries out, 'Experimenter, get me out of here! I won't be in the experiment any more! I refuse to go on!' Cries of this type continue with generally rising intensity, so that at 180 volts the victim cries out, 'I can't stand the pain', and by 270 volts his response to the shock is definitely an agonised scream. Throughout, he insists that he be let out of the experiment. At 300 volts the victim shouts in desperation that he will no longer provide answers to the memory test; and at 315 volts, after a violent scream, he reaffirms with vehemence that he is no longer a participant. From this point on, he provides no answers, but shrieks in agony whenever a shock is administered; this continues through 450 volts. Of course, many subjects will have broken off before this point.

A revised and stronger set of protests was used in all experiments outside the Proximity series. Naturally, new baseline measures were established for all comparisons using the new set of protests.

There is overwhelming evidence that the great majority of subjects, both obedient and defiant, accepted the victims' reactions as genuine. The evidence takes the form of: (*a*) tension created in the subjects (see discussion of tension); (*b*) scores on 'estimated pain' scales filled out by subjects immediately after the experiment; (*c*) subjects' account of their feelings in post-experimental interviews; and (*d*) quantifiable responses to questionnaires distributed to subjects several months after their participation in the experiments. . . .

(The procedure in all experimental conditions was to have the naïve subject announce the voltage level before administering each shock, so

that – independently of the victim's responses – he was continually reminded of delivering punishment of ever-increasing severity.)

6. Admittedly, the terms *proximity*, *immediacy*, *closeness* and *salience-of-the-victim* are used in a loose sense, and the experiments themselves represent a very coarse treatment of the variable. Further experiments are needed to refine the notion and tease out such diverse factors as spatial distance, visibility, audibility, barrier interposition, etc.

The Proximity and Touch-Proximity experiments were the only conditions where we were unable to use taped feedback from the victim. Instead, the victim was trained to respond in these conditions as he had in Experiment 2 (which employed taped feedback). Some improvement is possible here, for it should be technically feasible to do a proximity series using taped feedback.

7. The third condition also led to significantly lower obedience than this first situation, in which the experimenter was present, but it contains technical difficulties that require extensive discussion.

8. H. J. Laski, 'The Dangers of Obedience', *Harper's Monthly Magazine* (June 1929), pp. 1–10.

PART THREE

What Kinds of Problems People Face in Societies

Of all the problems which face mankind today, perhaps the greatest is the so-called 'population explosion'. Over large areas of the world populations are now growing faster than ever before. Often these areas are the poorest, either because they have few resources or lack the capital and knowledge to exploit them. Often too they have political and social systems that appear quite inappropriate to deal with the variety of problems associated with rapid population growth. In this situation, what can the social scientist offer? At first glance the answer might appear to be – not very much. In fact, however, the social scientist can offer a greater understanding of what is involved in the problems associated with population growth. He can define them more precisely, clarify the issues and so lay the foundations for more enlightened policy.

Some of the articles in this section deal directly with the current 'population explosion' in the less economically developed parts of the world. Others, however, have been chosen because they draw attention, if less directly, to crucial issues. For instance, the article by the demographer Norman Ryder deals with the rapid population growth in the United States and Canada during the 1940s and 1950s. It is included here, however, not so much for its substantive content, but because Ryder shows how important it is to refine the quantitative measures used to chart population changes if one is to talk at all sensibly about them. In looking, therefore, at the kinds of indices prepared by the United Nations, some of which are given below, one needs to bear Ryder's techniques in mind.

THE REPRODUCTIVE RENAISSANCE NORTH OF THE RIO GRANDE

*Norman B. Ryder**

TWENTY years ago it would have been simple to recount the fertility history of the United States and Canada. The annual number of births in the United States, for example, had risen to a peak of three million in the years immediately following World War I, and then, because the increase in the childbearing population was no longer sufficient to counterbalance the long-standing decrease in the average number of offspring, the reproductive output dropped slowly to a plateau of less than two and a half million during the depths of the depression. The immediate explanation of this fertility decline was clear: ever more couples were using birth control to achieve a family much smaller than they were capable of bearing. Then something happened. With the coming of war the birth tide again swelled to three million; shortly after the war's end the count reached four million, and it has continued climbing to the present record of four and a quarter million. From the lowest depression year until now, a mere twenty years, we have had an 80 per cent increase in births. Canadian procreativity has shown a similar history, although its level has been higher and there have been minor contrasts attributable to differing dates of entry into the two world wars and to differing times of response to the great depression.

The unravelling of the threads of explanation of this widely unpredicted phenomenon presents some perplexing problems. In the first place world demographic history provides no well-documented case of a rising trend in the fertility of any society and thus no analytical precedent. In the second place, although the post-war rise in the birth rate was shared by all industrial nations, only in the new countries – United States, Canada, Australia and New Zealand – has the recovery been sustained without diminution.[1] In the present essay, primary attention is focused on the United States. This emphasis is justified in part by the relative availability of the empirical data for adequate technical analysis and in part by the greater complexity of the U.S. fertility pattern, as will be seen subsequently.

Table 32.1 contains the information for a preliminary appraisal of the sources of increase in American births. The most recent quinquennium is compared with the lowest five years, 1933–7, of American reproduction. A 71 per cent increase in the annual num-

**Norman B. Ryder, Ph.D., Madison, Wisconsin, is a Professor in the Department of Sociology and Anthropology, University of Wisconsin. Formerly he was a demographer with the Scripps Foundation for Research in Population Problems, Miami University.*

ber of births between the two periods results from a 28 per cent increase in the population and a 34 per cent increase in the crude birth rate. Demographers have long been dissatisfied with the crude birth rate as a measure of fertility because its movements reflect the changing proportion of the population within the fertile-age span as well as modifications in reproductive behaviour. To achieve a measure free from this criticism, it has become conventional to calculate an individual birth rate for women in each reproductive age and then add up these rates, the sum being called the total fertility rate. This measure represents the mean number of births which a woman would have throughout her lifetime, assuming no mortality, if she experienced at each age the fertility observed during the period concerned. Since the increase in the total fertility rate has been markedly greater than the increase in the crude birth rate, we may infer that the gains in birth output have been achieved despite an unfavourable change in the age distribution. In fact, the increase in the number of potential parents during the period under scrutiny was only 10 per cent.[2] It is paradoxical that the influence of the age distribution on fertility should be negative during the baby boom. In the past, declines in fertility in every country were masked by a relative enlargement of the population of reproductive age. From 1920 to 1942 the American age distribution, as a factor in movements of the crude birth rate, was constant; but since then it has exerted a progressively depressing influence on the birth rate – precisely during the period of the rise of that rate.

TABLE 32.1
ELEMENTARY FERTILITY DATA, UNITED STATES

	1933–7	*1952–6*	*Per cent increase*
Annual births (in millions)	2·37	4·05	71
Population (in millions)	127·4	162·4	28
Crude birth rate (per thousand)	18·6	25·0	34
Total fertility rate (per woman)	2·27	3·53	56

TOTAL FERTILITY RATE AND REPRODUCTIVITY

The total fertility rate, which has risen by 56 per cent in less than two decades, has itself been weighed in the balance as a measure of fundamental reproductivity and been found wanting. The trouble is that it reflects two quite different elements simultaneously. One element is the average level of fertility actually achieved by the cohorts of women – that is, by women born during the same period – passing through the childbearing ages during the years in question; the other is the change that may be occurring in the age pattern of fertility from cohort to cohort. Such a change is implied in the idea of postponement of childbearing. The level of fertility in one year may be artificially depressed, and in a later year artifically inflated if the cohorts concerned decide to defer their childbearing to some extent from the earlier to the later year. In 1947, for example, the total fertility rate for native white females in the United States was 3·24. Analysing this total according to births by order of birth, one finds that there were 1·31 first births. This is impossible in reality because on the average women cannot have this many first births in a single year. Actually, spinsterhood and sterility set a maximum for this figure in the neighbourhood of 0·90. Generally speaking, the high fertility rates of the immediate post-war period reflected in good part the recovery of births previously postponed because of the vicissitudes of depression and war.

Enough time has now passed, however, to make postponement irrelevant for the analysis of present fertility. The maintenance – and even continued rise – of fertility has tempted some into the invalid position that the conventional total fertility rate can now be regarded as a fair representation of underlying cohort behaviour. The indefensibility of such a view can be shown again by breaking down the total fertility rates of the 1950s into the contributions made by births of different orders. The average numbers of first and second births are still far higher than a real group of women could possibly achieve. Since postponement is no longer involved, this suggests that another source of distortion is in the picture. There has been a progressive shift of the childbearing of successive cohorts away from the older and towards the younger ages. This development is inflating annual measures of fertility well above what they otherwise would be, to the order of 15 to 20 per cent. By using cohort fertility tables which have been prepared for native white females in the United States, one can compare the apparent rise in fertility, as given by the conventional total fertility rate, with the estimated rise in the fertility of temporally comparable

cohorts.[3] The former measure has risen by 51 per cent – from 2·26 in 1933–7 to 3·41 in 1952–6 – whereas the latter has risen by only 27 per cent – from 2·30 for the cohorts of 1907–11 to an estimated 2·93 for the cohorts of 1926–30.[4] Thus there has been an appreciable rise in cohort fertility, but it has been combined with almost as impressive an additional inflation of the total fertility rate – and hence of the annual number of births – attributable to a downward trend in the age of childbearing.

CHANGES IN FAMILY SIZE

Further insight into the character of recent fertility developments is provided by an inquiry into the location of changes by parity.[5] For this purpose we introduce the parity progression ratio, a measure of the proportion of women who, having achieved a given parity, proceed to have at least one more birth. Table 32.2 presents these ratios for the cohorts we have just been comparing and for the cohorts of 1890–4, selected because their mean parity (2·93) was the same as that estimated for current cohorts. The intervening experience follows a smooth path through the values

TABLE 32.2
PARITY PROGRESSION RATIOS OF U.S. NATIVE WHITE FEMALE COHORTS

Parity	*1890–4*	*1907–11*	*1926–30*
0	0·79	0·78	0·89
1	0·81	0·76	0·88
2	0·72	0·61	0·73
3	0·70	0·58	0·58
4+	0·69	0·63	0·52

shown. In parity-specific terms, the increase in cohort fertility between the depression and now has been the consequence of a rise in low-parity childbearing which has, for the time being, more than counterbalanced a continued decline in high-parity childbearing. Looking back to the 1890–4 cohorts, we see that the high-parity decline is a continuation of a well-established trend, but the low-parity rise constitutes a new element in American fertility. These movements, in combination, are reducing the variability of reproductive behaviour from one couple to another. There is an ever greater conformity to the small-family norm. The newest factor in this result is not the progressive elimination of large families, which is as old as our records, but the elimination of very small families – the childless and the one-child couples. There has been a modest rise in cohort fertility, and it is ample to forestall discussion of imminent population decline,[6] but the rise is exaggerated by the total fertility rate.

Our findings may be used as a basis for looking a few years into the future of United States fertility. In the first place, unless some development occurs which has not yet manifested itself in the data, the rise in cohort fertility is likely to cease soon because the sources of recent increase – the proportions of women in the lowest parities – have almost been exhausted. Secondly, current fertility – and thus the annual numbers of births as well – is unlikely to be buoyed up much longer by the changing age of reproduction. When the decline in the average age at which women bear their children is checked, as it eventually must be, then the conventional total fertility rate will lose its component of spurious upward distortion and consequently decline, provided cohort fertility remains constant. Finally, the proportion of the population within the fertile-age span will continue to diminish and thus tend to depress the birth rate until the large post-war cohorts reach the reproductive ages and thus reverse the age distribution within the fecund-age span. It is the writer's impression that, providing the politico-economic environment remains essentially stable, the conjuncture of these factors will produce a marked decline in the birth rate – perhaps by as much as 20 per cent – within the next decade. Should economic conditions deteriorate, a much more pessimistic forecast is in order because childbearing is now so largely voluntary.

CHANGES IN MARRIAGE PATTERNS IN U.S.

An important contribution to the recent transformation of the American fertility pattern has been made by changing nuptiality. Women today are more likely to marry, and to marry earlier, than at any time in at least the past two generations. It should, however, be noted immediately that this is not a post-war development, but merely the continuation of a well-established trend. Changes in the proportion who marry, however, can contribute in only a small way to an explanation of the decline in infertile

females. Even among the cohorts hit hardest by the depression, over 90 per cent of the women achieved marriage. But the decline in age of entry into marriage has been the primary agent in producing an earlier childbearing pattern, and this development, as explained above, has been a powerful, albeit transitory, contributor to the annual flow of births. This is not the place to attempt an explanation of the steady improvement in the marriage prospects of the young, but it may be worth while to sketch a few suggestions. First there has been an enlargement of the marriage market. Increased spatial and social mobility and the diminished variability of the population with respect to education, income and other social distinctions, have provided young girls with greater opportunities for meeting an acceptable mate. Another relatively unrecognised feature of American nuptiality which has depressed strongly the proportion of women who must remain spinsters is the high level of divorce. When a marriage is dissolved, the ex-husband and ex-wife are very likely to marry again, and about half of them marry single people who might otherwise have been left single. Finally it may be observed that, in a society like ours, the date of marriage is customarily determined by a compromise between romance and prudence. Since jobs are now readily available for the young man upon leaving school, and for the young woman as well, should the financial stability of the union require it, and since most couples are now equipped with the knowledge of and willingness to use effective means of fertility regulation if parenthood should for the moment seem a luxury, we should not be surprised at the contemporary early marriage age.

HYPOTHESES RE RISE IN MARITAL FERTILITY IN U.S.

There have been various hypotheses advanced to explain the recent augmentation of marital fertility in the United States. One suggestion is that it is now fashionable to have a large family. This is either a tautology, which presents us not with an answer but merely with another question, or else it is a denial of the possibility of understanding this strategic area of human behaviour. Another idea is that the insecurity of the outside world is impelling individuals to withdraw into the haven of fundamental values represented by the family institution. A further notion is that husbands with more time to spend at home are discovering – or rediscovering – within themselves an innate love of children.

Three reactions are provoked in the present author by these hypotheses, the list of which could be readily extended. One reaction is that the magnitude and even the character of what has happened seem to be misunderstood. An increasing tendency of couples to have more than one child has more than compensated for the continuing tendency not to have a fourth or fifth child. This is a more complex phenomenon than is suggested by the oft-reiterated statement that families are larger. A second reaction is the uncharitable suspicion that, given an opposite set of facts, equally plausible post-factum interpretations could be advanced. Finally, a third response is a scientific distrust of the persistent theme that the rise in fertility means a change in familial values, a new attitude towards childbearing, a reformation of reproductive institutions. In the writer's opinion, an idea like this, which may be true, should only be resorted to reluctantly and in the absence of plausible and testable alternatives because the area of inquiry is relatively impenetrable by present research means.

There is one line of inquiry which has received much less attention than it deserves, perhaps because it is obvious. Whatever desires for children couples may have must compete with other desires for the scarce resources of the couple's income. If that income rises, as it has in the United States in the past two decades, then the 'purchase' of children, like the acquisition of other consumer goods, may increase without altering the scale of values. People avoided having babies during the depression because they could not afford them; people are having babies now because they *can* afford them. The crude simplicity of this notion should not be taken as presumptive evidence of its invalidity. There is, admittedly, one large body of empirical data which seems to stand in the way – namely, the fact that for the same society through time, for different societies at the same time and for classes within a society, fertility has been correlated negatively with income. But a booby trap is concealed within this correlation. While fertility is being brought under control, the most important factor is the proportion of the group in question which is willing and able to control its reproduction. This proportion is directly related to income. But once a population, as for example the United States today, arrives at the stage when non-controllers form a negligible minority, the im-

portant determinant of changes in fertility comes to lie in the nature of goals and values affecting reproductive motivation. A mass of research findings has now been accumulated which attest to a direct relationship between income and planned family size. This meshes neatly with the fact that the sectors of the American population which have manifested the strongest procreative revival have been those most likely to control their childbearing. Also, the four countries which were cited earlier as having sustained a relatively high level of fertility in the most recent years are the world's wealthiest nations. In brief, it is suggested that the rise in cohort fertility is attributable to a rise in *per capita* real income among those groups which regulate the occurrence of birth. It need hardly be added that there are undoubtedly upper and lower bounds to this positive relationship between income and fertility.

DIFFERENCE IN PROCREATIVE HISTORY IN CANADA AND THE U.S.

Before bringing this account to a close, some attention must be directed to the respects in which Canadian and American experience has differed in the recent past. Canada has probably always had higher fertility than the United States. Principally for this reason, the Canadian age distribution during the inter-war era was not as favourable as the American to birth production. But the relative decline in the parental population which occurred as the United States moved into the post-war years was opposed by a simultaneous rise in this component of the Canadian age structure. The contrasting evolution is attributable in part to the earlier stage in demographic transition which Canada occupies and in part to a much heavier immigration, relatively speaking, of individuals of childbearing age. In consequence, while the parental population of the United States rose by only 10 per cent, that of Canada rose by nearly 40 per cent. The obverse of this observation is that age-specific fertility in Canada rose by only half as much as it did in the United States. The resulting convergence of Canadian and American fertility has been proceeding rapidly throughout the whole time of available fertility records. A detailed examination of cohort reproductive histories for Canada reveals, further, a pattern of change in the timing of fertility almost identical with that discussed above for the United States. The implication is that the rise in the total fertility rate (by about 30 per cent) for Canada in the past twenty years, can be explained almost entirely by the distortion of that measure arising from these changes in timing. Although the data on which these assertions are made are in detail inadequate for complete rigour and confidence, it is the writer's impression that whatever increases may have occurred in the fertility of Canadian cohorts have been trivial. Accordingly, in line with the analytical argument sketched above, the inference is drawn that in Canada the proportion of the population regulating fertility is still a more important variable than is the content of reproductive goals.

In conclusion, the writer has attempted to analyse the components of the reproductive renaissance in the United States and Canada. The role of increase in the numbers of potential parents has been characterised as unimportant in the United States but important in Canada. For both countries, a large proportion of the baby boom is attributable to a change in the age pattern of reproduction. In the United States, but probably not in Canada, a real increase in average parity has occurred. The structure of this increase, which is essential to its proper understanding, has been the virtual elimination of voluntary childless and one-child families. The hypothesis advanced to explain the convergence on the two- and three-child family size is that it is not the goals and values, but rather the means by which to achieve these, which have changed. From this may be drawn the inference that, given a high enough income level, the modern family will prove to be an adequate procreative institution within an urban-industrial environment.

Notes

1. F. W. Notestein, 'The Birth Rate Subsides in Western Europe', *Population Index*, vol. 21 (1955) 76–7.

2. This is the proportional increase in the ratio of the annual number of births to the total fertility rate. Although this ratio has no simple interpretation in and of itself, its movements reflect accurately the changes in the number of potential mothers.

3. P. K. Whelpton, *Cohort Fertility: Native White Women in the United States* (Princeton, 1954).

4. Obviously an estimate is required for the cohorts now in the middle childbearing ages because so much of their fertility has not yet happened, but a procedure developed by the writer and explained in a

forthcoming monograph permits an estimate based on continuation of the trends in the level and in the time-distribution of cohort fertility during the post-war period.

5. A woman's parity is the number of babies she has borne.

6. The present level of cohort fertility and mortality would support a growth rate of more than 1 per cent per annum in the stable population model.

POPULATION TRENDS IN DEVELOPING COUNTRIES

(from *Finance and Development*, vol. 6, no. 4 (I.M.F., Dec 1969))

THESE charts give a graphic impression of the basic demographic data underlying the World Bank's determination to help developing countries identify and find solutions to their population problems.

In a broad sense and for the longer run, the whole world has a 'population problem': space itself is a factor in the quality of life, even for the rich. But the Bank's most pressing concern is for the way in which unusually fast population growth distorts the age structure and diverts an excessive proportion of national savings away from productive investment in less developed countries, retarding their economic and social development.

The rates of population growth (fertility less mortality rates) depicted in Chart 4 are unique in history: man has never multiplied so fast. This results from progressively lower mortality rates, thanks to modern public health technology, while in most areas no comparable success has been achieved in reducing fertility. More and more countries are now attempting to apply available technology to the reduction of fertility with some promising results. At the present time, however, this appears to be making less progress than the control of mortality, and more countries will be confronted with a 'population problem' as their death rates decline; this will be true especially in countries with birth rates of more than 40 per thousand (see Chart 1).

The effects of this spread between fertility and death rates are now most notable where economic development has not progressed far enough to generate the automatic downward pressure on fertility which, historically in the more developed countries, has tended to accompany rising affluence. Chart 3 shows the impact on age distribution, with all its implications. As the chart indicates, more than 40 per cent of the population in South Asia, Africa and Latin America is under 15 years of age. In these circumstances there is a high ratio of dependants to the productive labour force and a rising proportion of childbearing women in the total population.

WORLD ACCORDING TO POPULATION AND BIRTH RATES BY REGIONS

Chart I

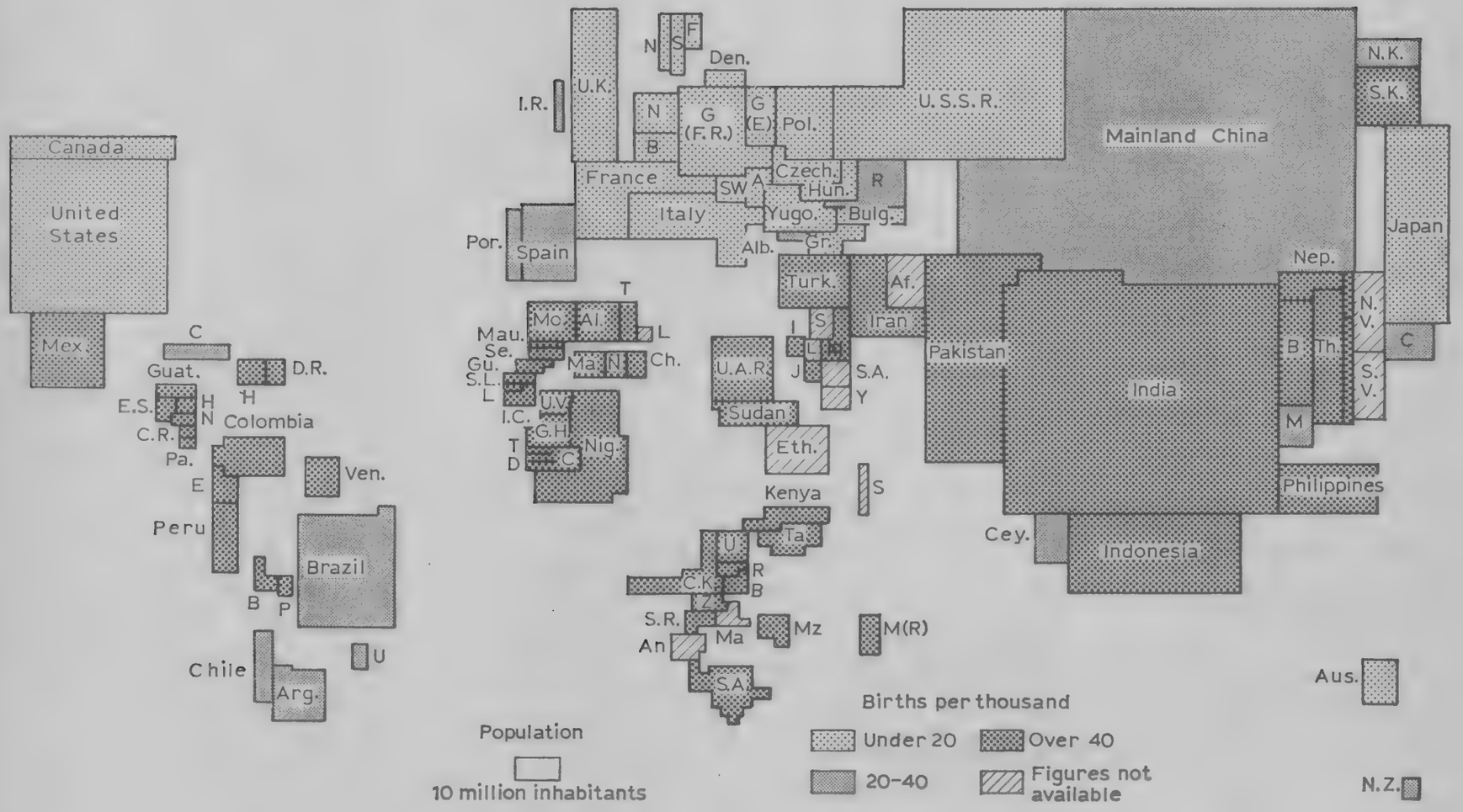


Chart 2

GROWTH IN WORLD POPULATION

Percent increase end 1967- end 1968

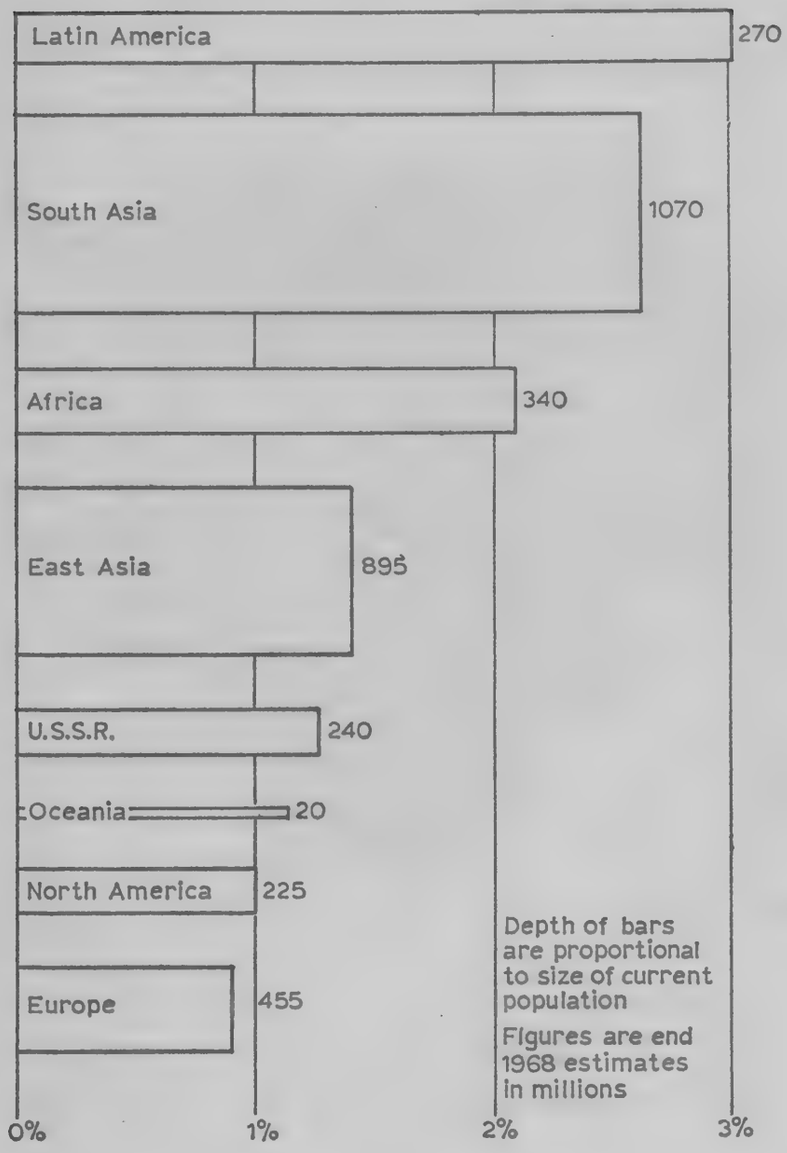


Chart 3

AGE STRUCTURE OF POPULATION IN SELECTED REGIONS

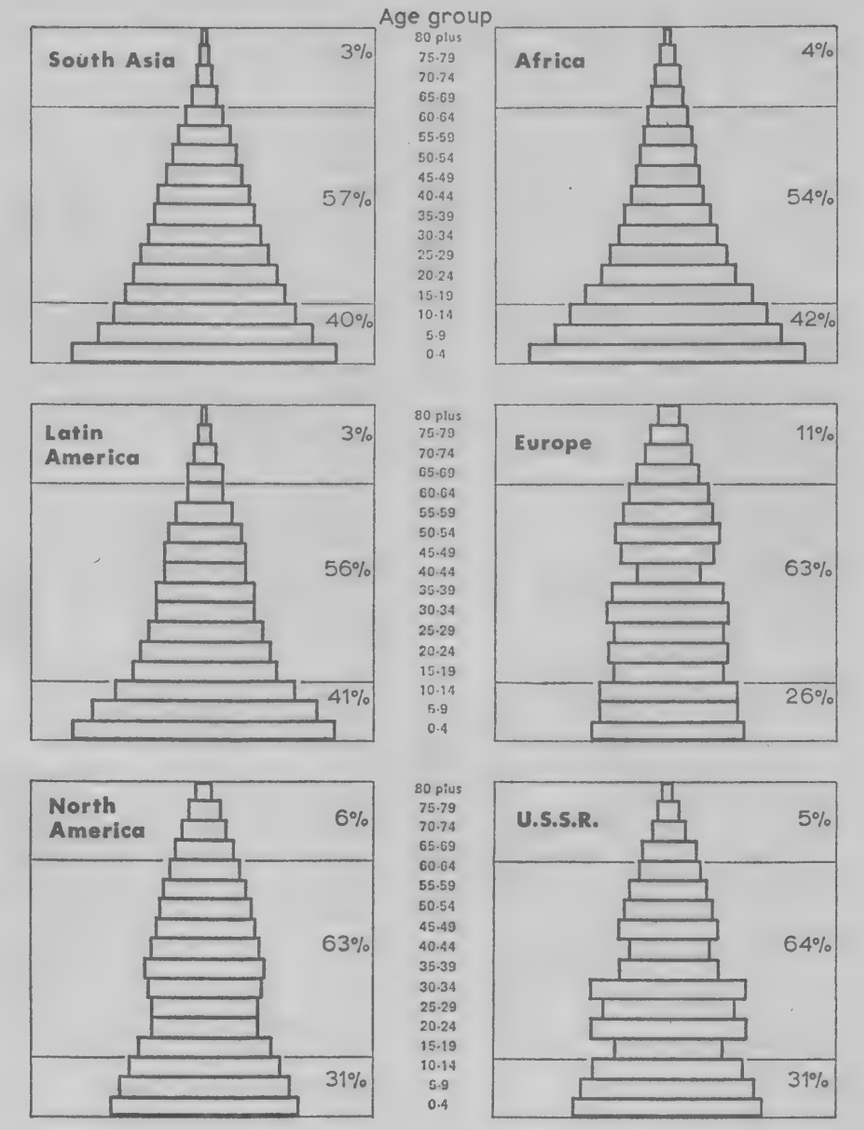


Chart 4

TRENDS OF BIRTHS, DEATH AND NATURAL INCREASE RATES IN SELECTED COUNTRIES

(Per thousand)

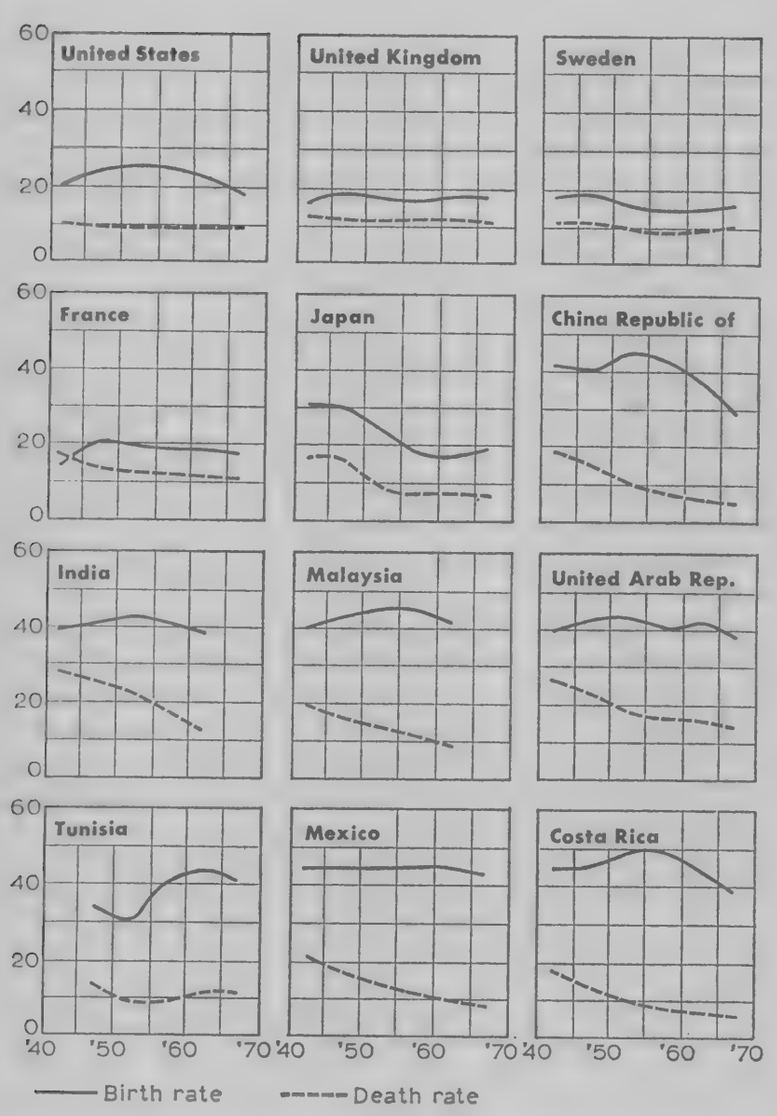
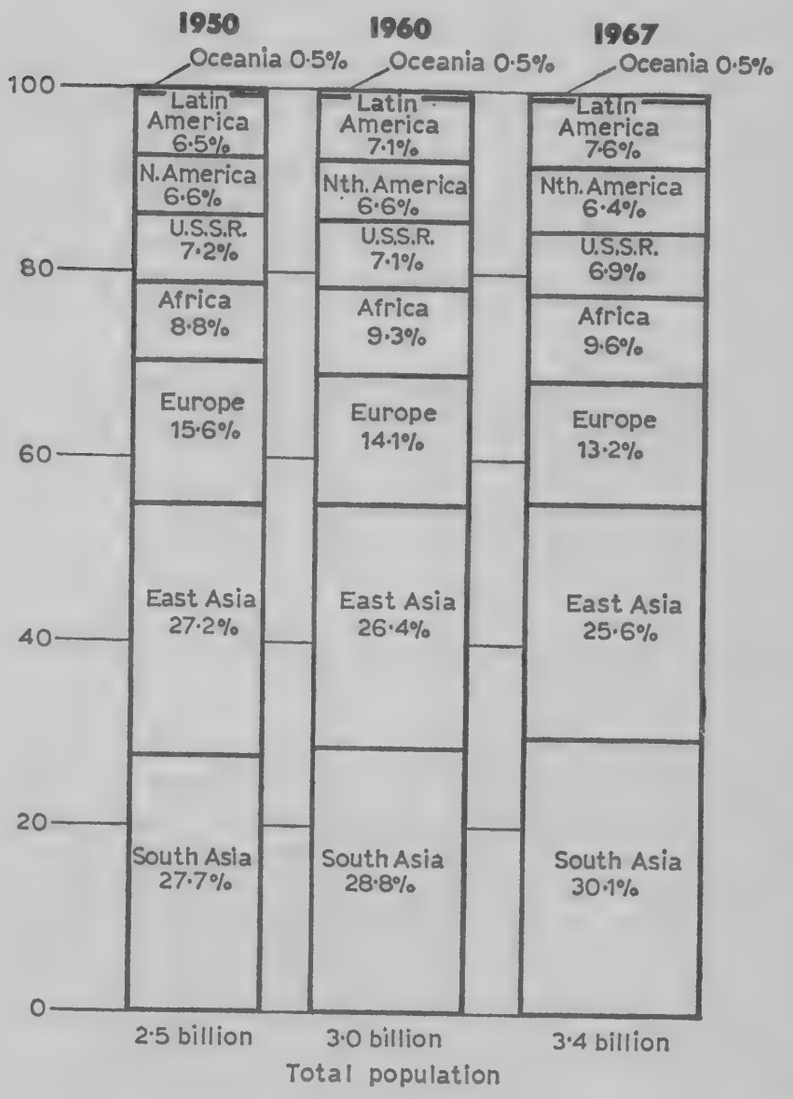


Chart 5

WORLD POPULATION BY MAJOR AREAS AND REGIONS



33 THE POPULATION GEOGRAPHY OF THE INDIAN SUBCONTINENT DURING AN ERA OF CHANGE

The inclusion of the article by Arthur Geddes ('The Social and Psychological Significance of Variability in Population Change, with Examples from India 1871–1941'), with its partial updating by Andrew Learmonth, serves several purposes. First, it counters the widespread view of the 'oneness' of the problem. Geddes, for instance, was able to show that India had well-marked demographic regions, and even though in recent decades, as Learmonth points out, a 'more blurred picture' has emerged, nevertheless the distinctions may well continue to have the social and psychological effects which Geddes noted and which need, therefore, to be borne in mind when trying to implement policies such as birth control. The article also demonstrates the value of one of the geographer's basic techniques – cartography.

THE SOCIAL AND PSYCHOLOGICAL SIGNIFICANCE OF VARIABILITY IN POPULATION CHANGE

with examples from India, 1871–1941

Arthur Geddes

I. THE PROBLEM; THE METHOD; AND SIX TYPES OF VARIABILITY IN REGIONAL POPULATION CHANGE

(*a*) *The problem*

For most of India's people, life is lived or death is found, close to the earth. When living among them and travelling from region to region one is struck by contrasts in the aspect and vitality of individuals and in the social life and the prevailing attitudes of the communities they compose. Marked physical and psychological contrasts are to be seen between the villagers of some tract where food and shelter are found in relative sufficiency and a fair level of health is enjoyed, and those of poverty-stricken tracts where the necessities of life are lacking. No less striking – though in a very different way – is the contrast of types found in the poverty-stricken class itself, as between the villagers typical of an arid 'famine tract' and those of an ill-drained, malarious tract in a moist delta. Between all of these people there are to be seen outward differences of physique and health, of carriage and physiognomy, together with other, more subtle, differences. These last are felt, rather than perceived, on first acquaintance; but one's impressions deepen with the passage of time and with growing knowledge of the characters of individuals and of personal and social relationships.

Conditions vary greatly not only from region to region, but also within the same region, from time to time. Thus, in a famine tract one saw and felt differences over as little as a year or two, as a crisis fell and passed; yet for year after year one saw little change in the lamentable aspect of villagers suffering from chronic endemic malaria, and reared in a region which had suffered since it was first struck by the disease, a couple of generations ago or more. As acquaintance with one's village neighbours grew, one's understanding was helped by the deeper knowledge of old residents and thoughtful natives of the region, and by the observations of discerning strangers.

Local and personal understanding, however intimate, is not enough by itself to carry one far towards increased comprehension of India and its people as a whole. The vastness of the country, the numbers of its communities, call for wider survey which should not lose sight of, but extend, what intimate observation can reveal. Now, central to all differing experiences suffered and undergone by individuals and communities is the incidence of survival or death, varying in different regions at the same or different times and periods. Primarily for this reason, one felt, the incidence and fluctuations in the balance of lives and deaths within a lifetime should be measured and indexed, and compared from place to place. By this means, one would be brought nearer to perceiving regional trends and discerning the rhythms of life and death, if indeed such were to be discerned. The method adopted is briefly explained below. When applied to the different Indian regions,

charted in time and mapped in space, the resulting indices proved to be demographically significant. Not only so, the resulting map of India – constituting, as it were, a regional record of major experiences and vicissitudes – was seen to form a starting point for a regional survey of social and psychological responses in Indian communities.

The 'net change' in numbers of population after ten, or after fifty years – up to the last census – was of course ready to hand for any Indian territorial unit, and was already a significant indication of recent conditions. For in India, where migration is a relatively minor factor, and where there is little check to population except want or famine or disease, the fact of bare survival is a sign that relative prosperity or good fortune has been enjoyed for a time. The unreliability of Indian vital statistics makes the decennial census all the more important; nevertheless such statistics may serve as a guide in intercensal periods. But 'net change' alone can conceal, as well as reveal, much that it is essential to know. Moreover (to mention only one statistical point), the element of chance in a figure of net change, which is no more than the difference between the populations enumerated at two arbitrary dates, will be much reduced for an index which takes account of enumerations, not at two, but at, say, six dates from first to last over a period of fifty years. Hence, one required an index of the sequence and character of changes over a considerable period. Such an index must evaluate the rising and falling waves which actually deviate from the mere 'average rate of change' deducible from the net difference between enumerations at two widely separated dates.

Such an index of deviations or 'variability' of population change should serve as a clue – crude perhaps, but fundamental – to the biography of a community. For this index concerns both individuals and the group or community to which they belong. Census statistics, by recording the enumeration of individuals by areas and period, sum up survival and mortality as totals, and form the basis for demography. But from the point of view of sociology or psychology the death of individuals is a matter of wider, or rather deeper, concern, for survivors are affected by the death of any member of their group. This is manifest when a woman is left a widow by the death of her husband, a child orphaned by the loss of a parent, or a mother bereaved by the loss of her child. Outside the family circle, though less immediately, the death of one person affects the neighbours, and his fellow workers, fellow villagers or fellow citizens.

To statistical sociology, in a word, deaths must concern survivors. Taken with the final result to date of the net change of population, the 'index of variability' which deals with intervening losses yields some account of the fundamental vital experiences of a community over a period, experiences whose results are both individual and social. Regional vicissitudes, due to extreme and widespread want or disease, not only cost so many individuals their lives, but have likewise caused the survivors to share want, hunger or sickness. And at the same time the impact of heavily increased mortality, with the large number of deaths occurring in families or wider communal groups, means the actual breaking of a still larger, though uncounted, number of human relationships.

The problem proceeding from these facts is to discover how far the variability and change of populations typical of the different regions of India may prove a guide to the regional character, rhythms and trends in social relationships; and to the prevailing states or conditions of feeling, mental life and social attitudes which result from the dynamic factors of regional population.

(*b*) *The method; and six regional types*

As already indicated, by 'net change' is simply meant the net difference, positive or negative (and expressed as a percentage), between the first and last of two or more census enumerations; but 'variability' requires a fuller definition. To estimate the variability, a smoothed curve (assumed to be exponential) was calculated, passing through the first and last available points, which were the populations of Indian regional units in 1881 and 1931. The deviations from this 'expected' curve, derived from the actual numbers at intervening censuses, were then summed, averaged, and expressed as a percentage of the mean population between the two dates. The greater the deviations, the greater their average expressed in the resulting percentage variability. This ranged from a minimum of 0·1 per cent to extremes of over 11 per cent which in one or two colony districts approached 20 per cent. One also noted the year (or years) of greatest deviation, with its negative or positive sign, the *minus* of disaster or the *plus* of recovery. In some provinces the census goes back to 1871, certain major facts being also known for earlier dates, while the last census

carries on our knowledge up to 1941. Thus, while the calculations dealt only with six censuses and a fifty-year period, as shown in the six graphs (Figs 33.1–6 on p. 629), conditions can be plotted, at least approximately, for seventy years. It need hardly be emphasised that statistics of health and of births and deaths are extremely inadequate in India; hence it is necessary to depend greatly upon the decennial census.

Let us briefly consider these differences of variability and net change of population, and the types into which they fall, in India during the half-century between 1881 and 1931. Where – so exceptionally for India without the help of immigration – population had virtually doubled in these fifty years, as in the annually flooded and silt-enriched eastern half of the deltaic Bengal plain, the steeply-rising curve of population was practically indistinguishable from the 'ideal' curve drawn to join the number first enumerated in 1881 to that of the last census in 1931. In this instance, increase was very high while variability was minimal. This may be termed high steady increase or simply 'high increase' forming Type 1 (Fig. 33.1). Where, on the other hand, endemic malaria had already taken hold of a population before the first census date, as in the western Delta of Bengal, the low net increase or arrest of population, recorded at the end of the period, actually corresponds to stagnation of the population throughout. Here the conditions of life from decade to decade, and even from year to year, have been almost as moribund as that of the rivers of this region which, in the peasants' own telling phrase, are 'dead' or 'dying'. The increase is low (or non-existent), and the variability is slight or minimal. 'Stagnation' may best describe this condition, forming Type 2 (Fig. 33.2).

For other regions, the change and 'variability' differed. Passing by intermediate examples of low increase but moderate variability, which may be described as 'intermediate' and classed as Type 3 (Fig. 33.3), let us consider the next well-marked type.

Where famine has repeatedly taken violent toll, as in the western Deccan or Rajputana, there are jagged graphs of population, in which relatively prolonged ascents are interrupted by the short, steep falls that tell of catastrophe. The increase is low or non-existent, but the variability is high to extreme. This is the type of 'recurrent crisis', Type 4 (Fig. 33.4). Before coming to the last type to be considered, we pass through examples of moderate increase and moderate variability, Type 5 (Fig. 33.5), which may simply be called 'Variable Increase'.

Finally, where, through the immigration of colonists, increase occurred as high as or even higher than that of Type 1 (perhaps to many times the initial population), the graph shows a definite change from the time colonisation began. The graph, flatter at first than the postulated curve, steepens sharply from the year (or census date) following the creation of a new and revolutionary factor. In an arid district this factor might be the extension of perennial irrigation to an area of steppes, which, as in the southern Punjab, brought forthwith an inflow of burly Sikhs, Jats and north-western Muslims. In upper Assam after 1911 the new factor was the commencement of mass immigration by sturdy Muslims from over-populated eastern Bengal. In such cases the actual curve taken is very different from the smooth 'expected' one; we find a notable and an extreme variability compared with the high to extreme increases; hence we know that this extreme variability marks, not a catastrophe but a demographic revolution, bringing in a new sequence of rising change. This may be described as 'immigration' or simply 'colonisation', Type 6 (Fig. 33.6).

While the calculations for the administrative districts and states of India – some 600 units in all – had been made and plotted in 1936, only a brief reference to these was published at the time; and I concluded their interpretation and publication only after revisiting India for a tour in 1938–9. The statistical results for the administrative areas were then published in a paper concerned with geographical distribution and essentials of causation. This was followed up by further environmental analysis and regional mapping. Statistical methods were next discussed, in the hope that others, better equipped statistically, may carry the method a stage further, in applying it to census results and vital statistics, in India and elsewhere.[1] The application of the 'index of variability' is also proving informative in tracing the historic contrast of demography, land use and culture to be seen between north and south of the central plateau in India, in two tracts which are similar in natural environment but utilised in a very different way by the Aryan-speaking inhabitants of the north and the Dravidian-speaking dwellers of the south. In addition, in certain notes attention has been drawn to specific environmental factors affecting these contrasts. Thus the present essay follows upon a series of

analyses of geographical environment, statistical method, and history; and the field has been prepared to describe, type by type, the social effects of the demographic change and variability. A summary of geographical methods must be given, to show the interpretation of statistics by topography.

Two maps accompany the graphs of the six selected administrative districts of provinces, or states, which are given in this paper (see Fig. 33.7 on p. 630). While the inset map shows the location of these six districts, the large map outlines the distribution of the six demographic types throughout India. Now the statistics for the whole country were compiled from administrative units such as the six districts here chosen to exemplify the 'types', each example chosen presenting homogeneity in the environmental conditions of human life throughout its area. But the boundaries of many other districts or units enclose territories each one of which may be highly diverse in physical environment and way of life. Yet an uncorrected administrative map assumes a uniformity of conditions throughout each unit, however diversified it may be in reality.

In such cases readjustment is possible with the help of geographical interpretation. In India, the human response to physical environment is topographically so direct, and so rapid in point of time, that when the chief environmental factors have once been broadly established, they can be applied even in the absence of detailed statistics: for example, in the relationship of drought to famine, of stagnation of water supply to endemic malaria, or of irrigation to colonisation. The real location of critical physical boundaries can be discovered and mapped by the geographer, principally by the help of detailed topographical study; and hence where the value of statistics has been vitiated by the arbitrary nature of administrative boundaries, corrections can be made and mapped.

To take an instance, the State of Bikaner is a census unit in which the northern margin of its arid steppes has latterly been reached by irrigation canals from the Sutlej river, perennially irrigated and colonised. The irrigated area of the state forms an isosceles triangle along the meridian of 74°E., with its apex at 30°N. and its base no more than 70 miles south at 29°N. (see Fig. 33·7); but the State extends further southward to 27°N. across 140 miles of dry steppe, forming part of the Thar desert. The boundary between these two tracts, which offer the greatest possible contrasts of landscape, can readily be drawn; and the geographer, by distinguishing the contrasted habitats of life, can determine the true areas of broad demographic homogeneity.

Thus geographical survey can form, as it were, a sociological and psychological reconnaissance. It may permit the sociologist who is about to attack a given problem, whether of fundamental research or practical 'reconstruction', to know where he must expect to find given phenomena or obtain support for his recommendations of appropriate measures of relief, re-education and reconstruction. The small inset map (p. 630) shows the location of the six regions described as examples, while the large map shows the results of the regional analysis published in 1942.

Lastly, with regard to the sociological approach, the suggestions made here as to the social significance of change and variability depend on the assumptions that there is an element of common experience, both within small communities – those of any family circle, caste, neighbourhood and village – and also from region to region throughout the country as a whole. The first, local, assumption may be readily granted. As to the second, it may safely be said that, while there are important regional differences in customs and religions in India, not only village economy but social customs have so much in common throughout the country as a whole, that the generalisations offered are not likely to be invalid. As to further questions raised in course of discussion, they are not to be accepted as assumptions, but are asked in the hope that they may be pursued in the course of social inquiries in different types of Indian communities. For example, one of the most important of these questions of social life which still await inquiry is raised in the discussion (section III (*a*)). This concerns the nature and consequences of differences in the relationship of a parent, and particularly a mother, to a child, under conditions of normal health and increase in a community, and under those of 'recurrent crisis' or endemic disease and 'stagnation'. In India – with the possible exception of trusted physicians and of missionaries in their own congregations – few men could attempt to investigate this question. There is therefore little mention of this in the examples given, which largely concern the life of men. Nevertheless, the question must be asked, and it is therefore raised, in the discussion.

Ignoring in this paper the intermediate types (Nos. 3 and 5) –

whose shades of difference I found highly informative and suggestive, as a geographer surveying the country as a whole – let us concentrate on one example of each of the four best-marked types: Nos. 1, 2, 4 and 6. For each type in turn, the people of a village and region which I came to know in India will be described, in order to show how the index of variability and change may reflect what struck me most in the condition of the people. We shall begin with 'recurrent crisis' (Type 4), as seen in a semi-arid tract of the Deccan lava region, and contrast it with 'stagnation' (Type 2), in the western half of Bengal. We shall follow with 'high natural increase' (Type 1), in eastern Bengal. The last examples will be taken from 'colonisation' (Type 6), as it has been seen among the colonists in north-eastern Assam, most of whom come from eastern Bengal, with mention of the more widely known conditions in the Punjab (from which the district taken as example is drawn). The chief points of physique and morale among peasants and labourers will be described for each example, and wider social and intellectual implications will also receive attention. Finally, it should be said that, while urban communities will be mentioned, the description in the following sections concentrates on rural life.

II. EXAMPLES OF THE FOUR CHIEF TYPES

(*a*) '*Recurrent crisis*' *seen in a Deccan* '*famine tract*' (*Type 4*)

When first I toured the north-western Deccan (of Bombay and Hyderabad) in 1921–2, the peasants were painfully recovering from the results of repeated droughts which had left them depleted of seed, cows and ploughing cattle, and had plunged them deeper into debt. They themselves were lean in body from inadequate food, and, but for government relief, there would, in fact, have been severe famine and a high death rate from hunger. Three years earlier, too, influenza had struck them down, leaving a corpse in perhaps every second family, or sometimes in every family, in a village. Famines had also occurred in earlier years; and the last great famine to go comparatively unrelieved, that of 1899–1900, had taken many thousands of lives (Type 4). Although I did not witness the passage of this 'pandemic' of influenza in 1918, yet on my first arrival in India it was so fresh in the minds of Indians and Europeans who had worked in some of the worst-hit regions, whether as doctors or missionaries, that their record of the people's terror, grief and despair was unforgettable. Influenza in turn had helped to bring about famine by causing a severe shortage of labour at the time when the autumn-sown crops should have been prepared for and sown, so that there was no reserve of food when the rains failed and crops were lost in the following monsoon. Thus misfortunes had accumulated and led to a major disaster.

A year or so after my first acquaintance with the Deccan peasantry, they had recovered considerably; and though they were still on the edge of want and as deeply indebted as ever, one sensed the difference in their ways. I recall the days spent around a Maratha village in 1922–3 with Maratha friends known and trusted by the villagers, and the nights spent on the *chowdi*, or communal verandah and meeting place, where we slept in the village. By day one saw and spoke with the men individually at work with their plough teams, or saw them strip to plunge for a swim in the cool waters of the tank-well at their midday rest. When night had fallen, there were long talks on harvests and failures, or on sowings with speculation upon the returns for which they hoped or the losses which they feared. The women one hardly saw, and could only guess their states of mind.

One day we had spoken to a peasant at his slow, laborious work of water-lifting with his oxen in order to irrigate his orange grove, his own and his animals' labour not then being required for the primary task of ploughing. That night he spoke of the high interest he had to pay on the loan he had raised to meet the cost of sinking a well in the rock, and the uncertainty of prices in the orange market on which he counted for repayment. Another peasant had decided to do like him, but spoke of the risk that even if a well were dug deep, it might not strike water; and then the land he mortgaged could never pass to his surviving son. With their talk was frankness and common sense, and some recovery of resolution, but not of confidence.

In spite of these troubles, the men gathered one night under a low moon, to dance their vigorous 'Morris' dances, as they might be called. They danced with the gaiety of recovered health, for their own pleasure, for ours, and also (no doubt) for that of their womenfolk, half-hidden in the shadows. By these dances – partly playful, partly martial – the youths learned light-hearted, but disciplined, control of rhythmic group movement, and the older husbandmen, weary from the plough, refreshed their limbs to

suppleness and revived their youth. Charades followed, played by boys with some elder as their producer, bringing peals of infectious, side-splitting laughter to us all. These people could once more face life with a sense of purpose, and even rise again to gaiety together. But behind them lay their memory of disaster, and the knowledge that sooner, or later, the rains would fail again, their crops be lost, and they, their children and their elders would face want, in a time of barren ground and famine.

While such impressions – as I believe – are true so far as they go, and are personally valuable to oneself despite their 'imponderability', we must ask how far they compare with the bare facts of the people's existence, vitality and mortality, their varying survival rates as measured by net change and variability. The net increase in this famine tract was generally rather low, having barely reached 30 per cent in fifty years. (For comparison the net increase of England and Wales in the same period, in spite of emigration and a falling birth rate, was 54 per cent.) Clearly, in a region where there is little or no permanent emigration and almost no voluntary check on population, so low an increase is a sign of a high death rate. The variability throughout the famine tract was very great, being 6 or 7 per cent or over, and was extreme in that particular district of this region which is later used to exemplify Type 4. Thus it becomes clear that while there have been periods of rapid increase, testifying to a minimum of fair health, these periods have repeatedly been interrupted by seasons or years of dire loss. Migrations, though appreciable, were a minor factor.

From the years and sign of greatest deviation one can particularise still further. By these we can tell, almost at a glance, whether those belonging to a given age-group at the end of a famine were at least allowed to reach adolescence or youth – although hardly ever their prime – in a period of normal achievement, of fair harvests, of relative sufficiency and advance; or whether, on the contrary, all of them knew fear of death by starvation and all to some extent shared the loss of friends, of at least one of their family, and perhaps of a parent. We can tell whether they passed through the successive stages of life with or without the personal, or communal, experience of loss, as child, as man or wife, as parent or elder. In India such generalisations can be made with relative assurance. Uncertain though we should be over prediction in successive Occidental generations of the proportion of married to unmarried, or the likelihood of parenthood in any age-group or class, in India the important experiences in the cycle of life are almost marked out at birth by the circumstances of the community. That the variability of population is so high, and the security of expectation so low, is the tragedy of an Indian 'famine tract'.

A regional description of peasant morale such as that just given needs to be complemented by some comments on life in the educated classes of the same region. In all of the four or five regions to be cited as examples, the intellectuals are drawn mainly from what a medieval European might have called the 'clerkly' castes and classes, and, to a lesser extent, from those he would have recognised as 'gentry' and 'merchants'. In India the keenness of mind and the energy of the Brahmins of the Maratha lands, the 'Marashtra', is outstanding. Of early Nationalist leaders, the names of Tilak and Gokhale will still be remembered by many European readers, as those of upright and vigorous men. In the history of the early nineteenth century, the prolonged resistance to the British forces in the Maratha Wars was due both to the martial qualities of the Maratha peasantry – Sivaji's 'mountain rats', whose sons have fought in both world wars – and to the organising leadership, civil and military, of the educated class. One may safely say that in spite of poverty, the *Chitpavin* or 'Poona' Brahmins continue to make their mark. Much the same is true of the mercantile castes and classes of the semi-arid western plateau. Gandhi himself, the son of an administrator in the Kathiawar peninsula of western India (20°N, 70°E), is of much the same Jain stock as the dominant broking caste of northern India, the Marwaris. Marwar (or Udaipur state, 24–28°N, 70–74°E), together with the rest of the unirrigated Rajputana, furnishes an extreme example of Type 4, 'recurrent crisis'.

(*b*) '*Stagnation*' *in Western Bengal* (*Type 2*)

In contrast to the high variability with rather low net increase in a 'famine tract', we noted the minimal variability with equally low (or still lower) increase – that is, stagnation or actual decline – in the area of 'dead' or 'dying' rivers in the west of the Bengal delta. This condition was particularly characteristic of the two deltaic districts of Jessore and Nadia which lie a little east of Calcutta and of the Hooghly and its uppermost reaches. Rather

similar conditions prevailed throughout the crowded alluvial rice-lands extending westwards to and beyond Patna (P on the map) – although this district itself, owing to the visitation of the plague, had a great variability and proves an exception among its neighbours. In the Bengal tract the physical ruin of the larger houses, built nearly a century ago, the growth of 'jungle' about the villages, and the weedy, ill-kept tanks, green with slime, prepare one for the condition of the people. The folk are haggard and lethargic; taught by experience, during the rains they stop work in a passing shower to hasten for shelter lest a chill should bring on fever and leave them trembling with ague; and although the grown folk are lean, the children are pot-bellied with the enlarged spleen of chronic malaria. In the absence of adequate medical statistics, one has to depend chiefly upon those of the Census. There is no doubt that in general, in this region, the high mortality reflects the incidence of malaria. The people's physical condition is paralleled all too closely by their morale. They are almost helpless in the grip of what has been called Bengal's 'malaria mentality' in which not the body alone but the very soul seems moribund.

What concerns us here is not simply the state of the individual in himself, or even the infection passed from the more to the less sick by the deadly mosquitoes in their faintly droning clouds. Invisible, but perhaps as infectious and as deadly as the mosquito or malarial parasite, there move the nebulae of mutual discouragement. Failures in mutual aid are reciprocated, as neighbours fall ill by turns; and chronically high death rates from infancy to age tell of bereavements that are almost habitual. When last I left Bengal, an attempt had begun to renew this western tract by calling in healthier peasants from the now crowded eastern tracts of the delta. It remains to be seen whether their labour will overcome the physical and morale problem of this situation, or whether they themselves will be overcome before the physical environment is improved. Be that as it may, here again the measure (of decline) and index (of minimal variability), plumbing these stagnant depths of mere existence, could have prepared one for the psychological atmosphere we felt about us.

(*c*) '*High natural increase*' *in Eastern Bengal* (*Type 1*)

It was after prolonged residence in parts of western Bengal – stricken and almost as desolate as those described, but to which, for all their sadness, I had become attached – that I first visited the eastern Delta. It was late in the rains – a period which in the western tracts is the peak of the malaria season. The floods were at their height, and the great rivers streamed like seas over the land, leaving the villages dotted like winding archipelagos among the waters. I shall never forget the sense of activity and vigour given by many of the riverside peasants and fishers. They moved freely in pouring rain, their dark, glistening limbs and torso stout and firm, ready with a ringing hail or even a shout of laughter. When the swinging rivers undermine a whole river-bank or sweep away an island sandbank in the estuaries and mouths of the Ganges–Brahmaputra, these peasants dismantle their neat huts, pack the sections on a little boat and, landing on a new sandbank, stake a new claim. There, they are ready to band together to work it, or to fight for it.

Poverty weighs heavily here too, and I would not wish to sketch a paradise, for the pressure of population had become extreme, although not to the point of suppressing effort. In a few years a plague of water weed, the 'lilac devil', had spread over the whole tract, clogging waterways and spoiling rice fields; but the people responded to guidance in fighting it, and they keep up the fight. As pressure has increased, peasants have flocked into the distant plains of upper Assam; and some, as has just been said, have even moved to the districts adjacent to the west, which were formerly the more densely crowded.

A sharper contrast between brothers in race and religion it would be hard to find than between the western and eastern Delta, between these peoples of almost neighbouring districts separated only by a narrow belt of transitional character. Once more a measure of the vitality of the people is offered by the dual indices of the net change in the eastern Delta (an increase of almost 100 per cent, doubling the population) and of the variability (its minimal percentage marking here the steadiness of increase). From the high vitality there follows the high frequency of unbroken circles of family and neighbourhood; while at the same time the index proves highly suggestive of the general outlook. It is all the sadder that during the last decade or two, local epidemics of malaria have continued to spread eastward, settling down as severe endemic malaria; and sadder still that famine struck the entire province in 1943.

Consider next the contrast in intellectual phase now existing between these last two types (Nos. 2 and 1), belonging to two tracts of one great cultural region, Bengal. Since the twelfth century, the intellect of Hinduism, and later of Islam also, has clustered along two principal zones: (1) on the west, the banks of the upper Hooghly (properly the Bhagirathi) and the red-soiled land to the west of it, forming the Rarh; and (2) on the east, the banks of the lower Ganges or Padma, round Dacca. By far the greater historic contributions to literature and learning came to birth in the western of these two zones, that which leads out to India's eastern coastlands; and the smaller, not unnaturally, arose in the more outlying, eastern zone. This regional predominance of the west continued through the British period almost up to 1870, when malaria broke out as a terrible epidemic, and then settled down in endemic form. I have described the effects upon the ordinary village folk of endemic malaria, seen at its worst, in parts of this zone, which were once a home of culture, arts and government. In matters of the mind the result of severe malaria was that culture wilted.

At this time those of the upper classes who could afford to do so fled to Calcutta, where their descendants have led in the professions; but few members of the upper castes born and reared in any village of the western zone have risen to eminence in this century. In the last century the eastern zone was slower in sending its sons afield to find new lustre, partly because it was more distant from Calcutta, and hence from opportunities in the modern world. Even so, in this century the eclipse of the westerners by the easterners from villages round Dacca was only possible owing to the ruin of health and morale in the west, and to the growing numbers, the relatively good health and the high morale found in the east of Bengal.

The general ability of the middle class or gentlefolk from the Dacca or Bikrampur area is well known in Bengal and eastern India; but in pursuing the matter I have been struck by the change, in the last two generations, in the personnel of the professions – of science, literature and art, as of administration and politics – which has followed the fading of the ancient west, the Rarh. As the earliest example of this historic western supremacy the name of Jayadeva, author of the 'Indian Song of Songs', may be cited to the Occidental reader; no such medieval classic was created in the Dacca region. At the end of the eighteenth century, Ram Mohan Ray was recognised in the Occident for his personality, as he is revered in Bengal for his insight. For the nineteenth century a list of other names could be supplied which, like Ray's, were all from the west, with one exception. Of the Tagore family, both branches are landed proprietors in the north of the Delta, and not strictly from either of these zones. In the twentieth century, however, although many men – like the distinguished Vice-Chancellor of Calcutta University, the late Sir Asutosh Mukerjee – might be of the western stock, they were rarely born in this old culture zone. In particular it is the region of Dacca which has sent forth the great majority of the staff of the Universities of Calcutta and Dacca alike, of the *swamis* of the Brotherhood of Ramakrishna (himself a native of Hooghly-side), of Rabindranath Tagore's 'University Group' at Santiniketan (on the Western Bengal uplands), or of the Bose Institute, Calcutta. A large number both of the officials of the provincial and indeed of the central government, and also of the Bengali leaders of Congress, are natives of eastern Bengal.

Now to turn back to our example of 'recurrent crisis' (Type 4), the Brahmins of Marashtra have held their own, in spite of poverty, where those of Western Bengal have lost their lead. By itself, the depressing effect of poverty on the literates in the last two generations might have been no more severe – perhaps less so – in the western half of Bengal than in the Maratha 'famine tracts'. But while famine is tragic at the time, only the poor suffer from its direst effects. With malaria it is otherwise. Such bodily infection cannot wholly be escaped by any class, even by those relatively protected by better food and absence of fatigue, and, for the last few years at least, by medicine and other preventive measures. Thus the morale of the whole community is directly sapped by malaria as it is not by famine. Such facts of causation and their effects cannot be clearly indicated by any one index of mere survival. Yet since no condition other than endemic malaria (or similar 'fevers') gives the combination of net arrest of population increase combined with minimal variability, this index gives at least a clue to the complex of physical and psychological states of a community over some two generations.

(*d*) '*Colonisation*' *in Assam and the Punjab* (*Type 6*)

In this, the last regional type, 'colonisation', by introducing a new factor, has revolutionised change and brought high increase with

high variability. We have noted the vigorous type of emigrant who leaves eastern Bengal for Assam. And the Punjabi is sufficiently outstanding, by his height and presence and – to those personally unfamiliar with him – by his place in literature, to need little description here. In the Canal Colonies of the Punjab the already self-respecting Punjabi holds his own. Brayne adds that the most startling progress in efficiency, honesty and cleanliness among the colonists come from some of the formerly disreputable gypsy cattle thieves of the Thar steppe who have been forced to settle down. This seems natural, in a place and time of well-founded, general hope and reasoned confidence in which, whatever the loss of his gypsy life may mean to a man, a wife and mother has everything to gain by acquiring a home; a home such as those she may well have slunk enviously by as a gypsy beggar, but can now call her children's and her own. Not every class will show so great a revolution in morale as do the gypsies; and we need not be surprised that it is among those who actually belonged to the environment now transformed, the steppe, that the revolution in outlook and social relations is either abortive or complete. Further north in the long-settled northern Punjab, though variability is often great and increase (locally) is low, both are due to the ready outlet of emigration, exceptional in India on so great a scale. This explains the apparent anomaly (for India) of comparatively progressive communities falling (in the statistical sense) into the 'intermediate' type (No. 3). Change in the prevailing ways of life, even in these northern parts, seems to be coming faster than in other tracts with equally long-established customs. Co-operation is relatively progressive, and the re-allotment and consolidation of fragmented holdings is increasingly undertaken by common consent of peasant land-holders.

Culturally, we find very different conditions in Assam and in the Punjab. Assam is a land of ancient Hindu culture which is being 'colonised' by a Muslim peasantry. If these incoming Muslims take to modern instruction, as they ultimately will, it is difficult to see how Assamese culture can remain uninfluenced. Meantime, however, the Bengali peasantry by its vigour is adding to the economic resources of the province, although their presence, and the cultural problem they offer, can hardly be said at the moment to have reached beyond the early stage of arousing defensive attitudes in the intellectual life of Assamese Hinduism.

In the Punjab, the actual Canal Colonies do not straightway attract the professions, though merchants flock thither as brokers of the cash crops grown there, and as money-lenders. The professional classes are to a great extent enriched at a distance; from 1921 to 1931 Lahore grew faster than any other Indian city and has maintained this rapid growth up to 1941. Lahore's University is an active centre, the city's trade and its administration are busy; Amritsar and other old centres have their colleges too. It is the progressive outlook of the Canal Colonies which quickens the life of the province as a whole, and invigorates that of the educated classes.

The 'regional types of variability-change' may be traced over the maps. In Bengal the 'high steady increase' (1) of the east contrasted with the 'stagnation' (2) of the west. These types give place in Assam to 'variable increase' and 'colonising immigration' (5 and 6). Westwards, 'relative stagnation' dominated the crowded rice plains to the floodplain of the Gogra, with only two exceptional districts, of which Patna (P, at the Sone-fan) suffered from plague, and another from flood, owing to its situation at the Gandak–Ganges confluence. At the Sarda cone and its skirts variability was greater (Type 3, 'intermediate' with local 'crisis'). 'Stagnation' (2) continued west along the plateau edge, recurring towards the Himalayas and in the malarious W. Jumna Canal tract (lying to N. or S. of Delhi, D). Still farther west to the Chenab and Jhelum, with increasing variability where irrigation has not reached, 'intermediate' conditions (3) alternated with 'recurrent crisis' (4). Where irrigation has added to productivity these gave place south-westwards from Lahore (2) to 'high, steady increase' (1), to 'variable increase' (5) and to 'colonising immigration' (Types 1, 5 and 6, in successive belts). In Sind 'variable increase' prevailed on the eastern bank of the Indus, but 'intermediate' conditions marked the western bank until the irregularities of inundation were replaced by perennial irrigation. Around the mountain fringe of British India one is reduced to guesswork, and 'variable increase' is suggested as a prevailing type.

To west and south of the Punjab plains, beyond the irrigation zone, 'crisis' recurred, from across the Thar desert by Ajmer (Aj), through the Malwa plateau to southern Rajputana, to the rivers Narbada and Tapti (21°N). With a break at the Tapti basin, 'recurrent crisis' also characterised the semi-arid western Deccan (in the lee of the Ghats), southward to 15°N. Around Nagpur (N), in the better-watered Central Provinces (20–24°N), 'intermediate' conditions prevailed, moister or more wooded parts showing a higher increase (1). In much of the eastern plateau from the

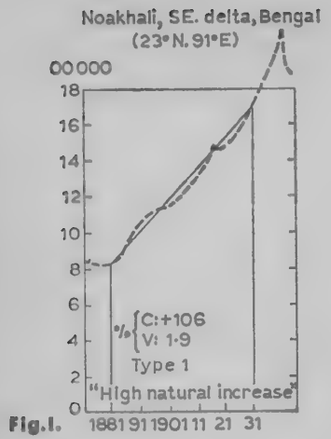


Fig.1.

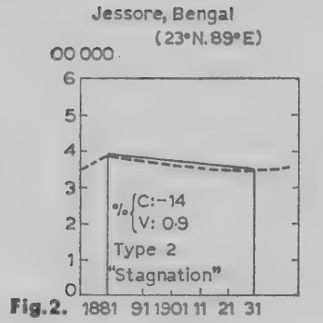


Fig.2.

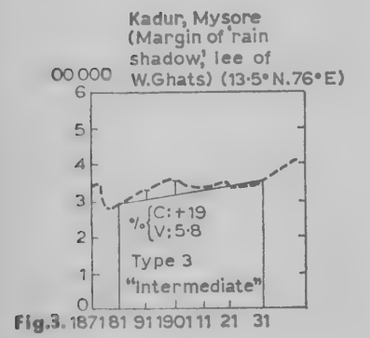


Fig.3.

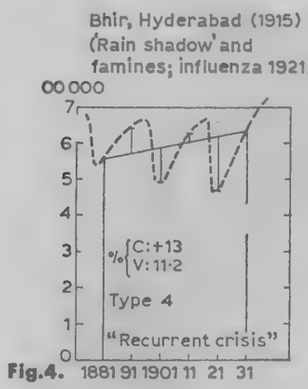


Fig.4.

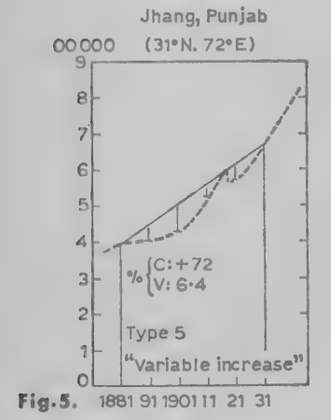


Fig.5.

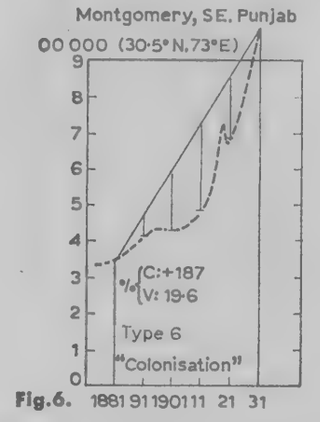


Fig.6.

Fig. 33.1–6 Population change and variability, six types. (In the internal legends of these figures, C = net change and V = variability (see p. 621–2). The vertical ordinate of the graphs is population.)

basin of the Mahanadi to Hyderabad (4) and the Kistna, the 'high' but usually 'variable' increases gave type 5, and where mining or industry have added their quota, there was a 'steady, high increase' (1).

In the lands of Telugu speech south of the Kistna, drought and famine accounted for the recurrence of 'crisis' (4), shading into (3). But on the heights of the Mysore plateau round Bangalore (B), fairly steady recovery from the last great famine (1880) gave Types 5 and even 1. Most of these tracts lay on the marginal line of more than average increase, 50 per cent, even where 1 gives place to Types 2 and 3, in the rain shadow of Ceylon. Along the west coast, 'intermediate' conditions prevailed from the Kathiawar Peninsula (22°) and Ahmadabad (A) southwards past Bombay city through the Konkan rice tracts to Goa (15°N). In the more steadily watered southern portion with its widening strip of alluvium, 'steady' conditions with only 'moderate increase' (2) gave place to 'high increase' (1), culminating in the southernmost state, Travancore.

Ceylon's increase was 'steady and high' as a whole, but regionally 'relative stagnation' prevailed on the sand-bar coastlands and lateritic uplands in the lee of the monsoon, to the north, while the plantations of tea (and coffee) led to 'colonising immigration' by Madrasis, though this, too, was steady during the period. The rice, coconuts and fish of the coastlands have nourished steady growth, and Colombo's population was tripled. Burma and Ceylon are physically comparable in being for the most part well watered, though with 'dry' belts in the north, and to this the high to extreme increase of population they share is greatly due. In deltaic lower Burma and the ridged coastlands are found conditions common to eastern Bengal and Travancore and their hilly fringe. 'High, steady increase' (1) therefore marked the longer-settled parts of this great tract. Around them, 'colonising immigration' (6) reclaimed the tidal jungles for paddy and the surrounding ridged uplands for plantations, chiefly of rubber, while hill peoples enlarged the upland clearings for upland crops, probably with 'variable increase'. In all the regions predominantly peopled by Burmans (as distinguished from Shans and others) those native to the country (enumerated as Buddhists) are stated to have increased by over 75 per cent since 1901, and by 100 per cent or over since 1881. Even after making allowances for errors, we may say the truly native population of Lower Burma must have doubled in the half-century. In addition, immigration from India contributed to the increase.

For Upper Burma, from what is known of conditions previous to the northward extension of the census in 1901, it seems clear that in certain areas of the 'dry belt' towards Mandalay (M, 22°N) the hardships suffered, which led to some emigration to the delta, constituted 'intermediate' condi-

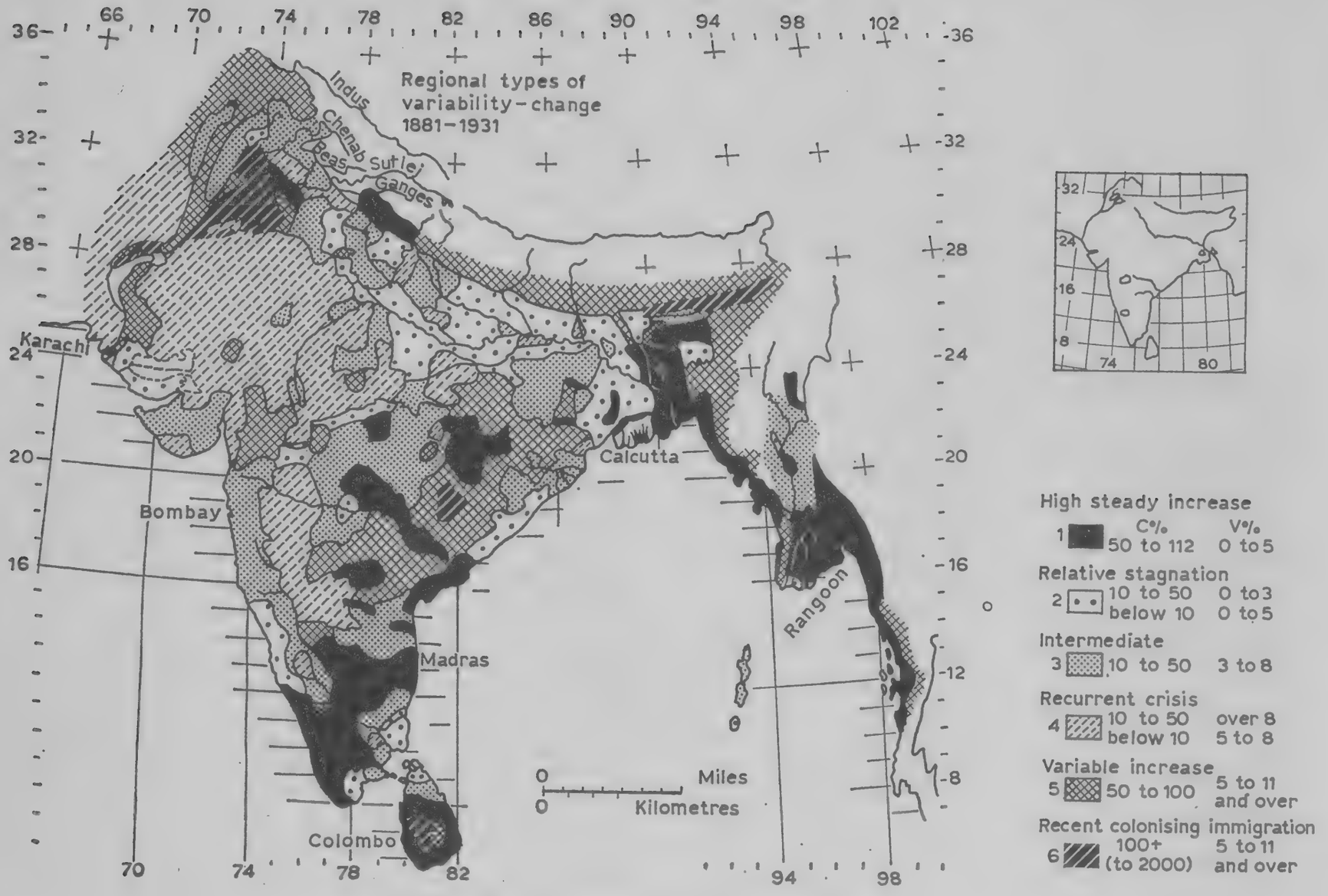


Fig. 33.7 Regional types of variability change

tions (3). Locally the development of oil led to 'high, steady increase'. But for the disasters of war, recent extensions of irrigation should have done much to extend the protection of agriculture and population around Mandalay.

III. DISCUSSION

(a) Parent and child, the response to bereavement; and the observer

No one who has passed afoot through Indian hamlets can forget the appeals made to him to save a sick child or relative, nor the suffering graven on the faces of watchers about a sick-bed. Yet there is so much bodily sickness and mental suffering on every hand, that none but one quite sheltered from birth, then suddenly let free in village or town, could mark every sick and maimed or unhappy fellow-creature. For in India any one of them may be a spectacle such as seared the eyes of Prince Siddartha when first he rode abroad, and may ring the heart and conscience of any follower of the Buddha. The Hindu doctrine of Karma too easily allows of passivity in the face of other's suffering, as due to divine law; the Muslim, too, is taught to accept that whatever happens is the will of God. But in spite of the rigid separatism of caste in Hinduism, India has given the world many of our earliest and greatest teachers of pity, consolation and brotherhood. Yet most of us learn to pass many a sufferer by, the very frequency of pain and loss making it inevitable and probably necessary for sensitivity to become calloused, even if the heart be not hardened within. For this reason, the perception of pain and the judgement of either a native or a stranger is unlikely to be warped by undue 'sentiment'; rather, indeed, the reverse. In any case, mental suffering from loss is imponderable; all that can be affirmed is its widespread occurrence. What can be measured is the variable incidence of its prime cause, mortality, and, indirectly, the bereavement which follows.

In regions and cities where child mortality is especially severe the effects upon women of the loss of their children must be of deep significance in morale. When I lived for a time in the poorer quarters of Bombay, the city's infant mortality exceeded five hundred per thousand; every second mother among one's neighbours lost her child in its first months. In his own western Bengal, Tagore knew and portrayed a mother's feelings for her child sickening of fever. The effects of such losses sink deep. The task of determining how far mortality among children contributes to morbidity and mortality among mothers who have actually suffered loss may lie as yet outside the scope of formal medicine, although the recovery from losses of this kind has long formed a central theme of psychoanalytic research. But some influence of such a loss on the mother has been noted by every observer, be he a sensitive Bengali bard or perceptive diagnostician.

As the regional incidence of loss varies, so must the range of its effect upon mothers. A mother may be wounded to the heart, whether an epidemic has struck down her child almost with the suddenness of Herod's massacre, or whether she has suffered the fretting as sustained anxiety brought by endemic disease, from which she too may never be free. But the effect on her attitude to life may differ according to the nature of the wound, depending on whether it was a sharp stab or some slow lingering cause. The first, by its shock, may be the more severe at the time, yet the second be the more wearing in the end.

A question as to the effects upon parents, particularly mothers, of the loss of their children, was asked of a European physician, an officer of public health. Speaking from a lifetime's struggle with disease in India, he questioned in turn, 'Are not the mass of the Indian people so accustomed to death and loss as to have become inured to them, inured to the point of fatalism?' We must admit this to be more than true wherever disaster has been widespread: sensitivity becomes blunted in self-protection as a repeatedly wounded area may heal with a hard protective scar. The doctrines of Karma and of 'Kismet', reinforcing this, may reflect an age-old self-defence. But the very statement that disasters breed fatalism or resignation is no more than another and perhaps stronger way of saying that disastrous experiences, whether in the individual or through a whole generation or more, have a deep effect on human beings, and lead to a protective resignation which inhibits both painful feeling and the confidence needed for constructive action in many spheres of human life. We may say, in other words, that the change and variability of a regional population must provide some guide to the probable state of mind of that population. But it is perhaps useful to distinguish between outward resignation, or conscious fatalism, and those deep and often unrecognised attitudes which may, nevertheless, greatly affect the quality of human relations, particularly within a family. Those closest to a mother who has suffered loss suffer with her in some degree, but she who is inconsolable has herself little power to bring

comfort, be it to her husband or his parents, or to her own remaining children.

Beyond the initial clue to experience furnished by variability, the unrecorded complexities of response to death and disaster are great; and the detailed differences to be found must be infinite. As already said, it is not only the sensitivity of the stranger which may become hardened or calloused to suffering, but that of the people themselves. On the other hand, speaking of both villages and towns in India, I would say that almost everywhere one is aware of the presence, or rather the predominance, of feelings a European naturally shares, despite the difference in his cultural background and experience. Equally, he cannot fail to note relationships and responses akin to those to which he is accustomed in Occidental cultures, among friends or neighbours of any 'class', in comparable circumstances of peace or war. In rural parts of India, on the other hand, a rare exception is furnished by castes who have adopted a criminal way of life.

It is in the towns and cities of India (as in those of Europe) that one most frequently meets with evidence of that callous indifference which psychological investigation so regularly displays as an entirely logical – though unconsciously determined – protective device in those whose social apprenticeship in the family offered no hope of security, of affection or indeed of anything except a bare existence. This is truest of huge towns such as Bombay or Calcutta. Under the full impact of the commercial and industrial revolution these cities grew up almost unchecked by legislation of the kind which began to abate the horror of British ports, manufacturing towns and mines, more than a century ago. In walking day after day through the streets, alleys and courts of a town, in order to prepare a 'diagnostic survey' preparatory to replanning, one gathers a considerable quantity of observations. The multitude of little incidents and impressions of individuals which accumulate in the memory, ultimately fuse into generalisations upon the characteristics of particular socio-economic groups. Such generalisations take form all the more readily and reliably because the members of these groups are recognisable by colour, features and dress, and dwell together in neighbourhoods and quarters according to the prevailing origins and occupations which determine their traditional caste or class.

In certain of these groups – not necessarily the poorest – one is made sharply aware how low the level of parental feeling can sink. The aspect of the men is repulsively greedy or brutal, and the rate of venereal disease is visibly high, among both sexes. The despair, frustration and anger which lie behind cruelty, delinquency and loveless promiscuity are only too obvious in these surroundings. Many of the women appear indifferent to their unwanted brats, begotten by lust out of lechery, children whose death may bring the parents nothing but relief by easing their burden, so long as one or two survive to furnish – however grudgingly – a means to provide for their parents' old age. Of Bombay and its infant mortality my impressions are of two very different sorts. When I recall a certain windowless, rat-ridden room, shared by two families, in a corner of which lay a child in the last stages of fever, I remember the faint glimmer of a question in its mother's eyes before they dropped again in the drawn sorrow of her watch. I also remember, in a quarter where I shared for a time a room with Indian student friends, the face of many a woman who had so clearly been forced to strangle affection in her heart. Prostitution was common in that quarter, and the infant mortality which had prevailed there for some years past, a mortality of more than one in two, was not all accompanied by mourning. The very conditions precluded such a possibility.

Apart from conditions of disrupted society such as those of an old and festering town, or a modern industrial port, other factors of an ancient origin must be remembered. If it be difficult for one native-born to unravel the mingled strands of feeling in peasant parents bereaved of their children, it is almost impossible for a stranger to do so. The observer's task is made the harder because custom may prescribe stated times or occasions for the expression, exhibition or concealment of grief. The observers' sympathies or prejudices must also be remembered. Among British witnesses, the differences in customary expression of concealment of emotion are considerable, according to country and class or origin, and their view of Indian customs may be coloured accordingly. An Irish witness, if accustomed to distinguish among the peasantry of Galway or Donegal the traditional artistic expression of genuine emotion from a counterfeit or sham, is prepared in India to discern the oratory of heartfelt mourning from either coldly ceremonial expression or hysterical exhibitionism. Another witness of the same scene, if moulded by English 'public school form' and accustomed

to value – perhaps to overvalue – reticence and self-control, might lump together the true with the false.

Beneath the veil thrown by established forms of expression or reticence lies yet another screen. The direct and spontaneous reaction of a wife or mother to loss may be complicated by prescribed social relationships, particularly those with her parents-in-law, with whom she may be dwelling. Her own primary human response – whatever it be – will be overlaid by custom, and her thoughts diverted to her place and her part in this. The primary grief felt by a father who has lost an only son may come to be overborne and almost replaced by the consciousness of religious stigma and loss of social credit. But social customs may aggravate 'natural' grief rather than alleviate it. A sorrowing woman may be blamed for the death of husband or child, or for her barrenness, the result, it may be, of premature child-bearing. The corpse of a wife who has died before her husband – worn out perhaps before her time – is given full honours; but widowhood among many Indian castes is tantamount to a kind of disgrace. Equally, custom may not be satisfied though a wife produces children if they are not sons but merely daughters. As in communities in other lands in which custom has caked hard over spontaneous feeling, so in these the crust of social prescription must not be forgotten. In spite of all complexities, however, some impact of loss is inescapable. My own impression – for what it may be worth – is that the influence of custom may be overestimated and that of primary feeling underestimated. I have been struck by the difference in atmosphere in a normal family from that found on entering a peasant farmer's household where a widowed husband had called in a sister of his own to mother his children, his sister being herself a 'widow', whether from betrothal in childhood or from actual marriage in later life. In most Indian villages a father is as ready to take part in gentle, playful care of his children as in any other countryside I know; but the lack of healthy children in so many regions saddens the play of expression in the villagers' faces and dims, or quenches, the light in their eyes.

(*b*) *Historical geography and contemporary morale in peace and war (1931–47)*

A further point requiring attention is the period of past time over which an index such as we are now considering may throw light. From what has earlier been said, it will have been realised that the 'famine tracts' have always been liable to drought and its effects, and that the great variability in population over the last fifty years forms a sample of ancient conditions, much attenuated in severity since 1900. Indices of relatively recent situations may also be suggestive of historic periods and changes. In the Deltaic regions cited, history makes it clear that malaria, relatively harmless for generations in the past, suddenly fell upon a subregion of a great plain, ruining health and morale for generations to come. Thus both for arid 'famine tracts' of 'recurrent crisis' and for humid deltas whether of 'high natural increase' or 'stagnation', the change and variability of population throws light on the demographic conditions of the earlier past; and they should, therefore, help us to understand the social conditions under which regional forms of belief and precept, of doctrine, symbol and custom, took their shape, affecting mind and morale today and for a long time to come.

As to the future in the different regions, in the areas of colonisation there may exist the possibility of a rejuvenation alike of Hindu and of Muslim. In our period the peasantry there are gaining access to education not only through print. Among the Punjab peoples, military and naval training (for all its drawbacks) showed signs of having a stimulating effect; and so too has the experience of war, in spite of war's drain of picked manhood. On leaving the Indian Army, servicemen were encouraged to settle as colonists, and could hope to command relative well-being in civil life. The 'golden age' of Hindu tradition and history was one of internal as well as external colonisation, a time when fresh land invited advances, clearage and tillage, and called forth religious vision and fresh efforts of social organisation. Of Islam it might be said that from the caravan came the Koran, and that the Koran was followed by conquest and expansion. The physical experience of expansion to new lands is today realised by colonisation in the Punjab and Sind, and in Assam. With this has come a liberation from the social burden of a crushing death rate, which crushes whether it strikes by 'epidemic' or 'endemic' incidence, through high variability or low. With liberation from this burden, latent possibilities of advance in morale and in cultural level will at least be open.

While the generalisations just outlined apply to times of peace,

other problems relevant to our theme have faced – and will face – India in times of tension and of war. Among them is the problem of raising national military forces, in which there is a tangle of factors other than those of physique, intelligence and a tradition of courage. There are, for example, the basic considerations of political allegiance; the development of martial traditions and habits in some regions and their virtual absence or extinction in others; the varied reasons for the choice of military service by, and of, certain classes and castes to the exclusion of others; together with regional, religious codes or recent ideals of conduct, from holy war to non-resistance, in which other aims than defence against an immediate aggressor are concerned. Thus no simple correlation can be made of the regional types indexed, with the attitude towards 'military service' and military potentialities. It may, however, illuminate our theme if we make a few comparisons, along the north from the Punjab to Assam, of past attitudes and experiences in these matters.

Taking the regions which sent forth the highest proportions of recruits to the combatant forces in the years up to 1946 a few correlations stand out (see Fig. 33.7). The wide area of the Punjab plains under Type 6 is noticeable along with areas of Types 5 and 1. Even areas of Type 3 and 4 so closely resemble these others, that one can perceive that in spite of great poverty their people have found some relief from scarcity and threatening famines, thought not from epidemics, by emigration to the Colony tracts. The virtual absence of Type 2, 'stagnation', is noticeable, except beside the river Jumna, near the eastern border, which supplied few combatant soldiers. East of the Jumna, in the Himalayan tracts of the United Provinces, Type 1 corresponds to an area in which high-caste recruits were found. (Statistics for the Punjab Himalayas and Kashmir are less reliable, and incomplete.) Farther south extend the lands of 'recurrent crisis'. Although, as one would expect, the physique is poorer here than in the Punjab, this tract furnished soldiers to the states of Rajputana and, farther south, to the Maratha regiments. The plain of the Western United Provinces, with Types 4, 3 and 2, furnished some recruits of high caste. Farther east, however, comes the area predominantly of Type 2, 'stagnation'. Although comparatively sturdy workers are to be found there, and although what now forms Bihar furnished high-caste Sepoys to the East India Company's successful 'Bengal Army' until the mid-nineteenth century, it is hard to estimate the physique and combatant potentialities of these castes as compared to the qualities of the men of the Frontier and the Punjab. It should be remembered that it was not lack of courage in the first place which deprived the Biharis of the opportunity, or burden, of military experience. Of the defects of the pre-Mutiny armies, wrote Sir John Lawrence, 'unquestionably the worst, and the one which operated most fatally against us, was the brotherhood or homogeneity of the Bengal Army'. This suggests that, given their country to defend, there are Biharis who, well led, could do so worthily. Military skill, however, is another matter, demanding time. Of the Hindu Hill Dogras, Sir George MacMunn has written: 'Faithful, enthusiastic, honourable, unaccustomed to stealth, in warfare they have to be trained to take cover.' In this way they contrast with the wild Muslim Hillmen of the Afghanistan frontier who are only too accustomed to use stealth to kill an enemy.

When we enter western and central Bengal, where Type 2 is of almost unmitigated extent and fixation, the endemicity of malaria is so severe as to have made the search for soldiers all but vain. One has to have seen the effects to realise how paralysing this disease can be, from youth upwards. On one occasion some lads of Rabindranath Tagore's school, whom I had been asked to take out, accompanied me on a short exploration of their own countryside. Not only were they quite well fed and living on a fairly healthy upland, but they regularly played vigorous football. They followed me for three hours, returning a little late for the evening meal but in excellent spirits. The elderly headmaster, no athlete, but a man whose mind and scholarship was held in respect, was deeply concerned at my having taken them out so long, and upbraided me, without due cause that I could see. But next morning, whereas the lads in a malaria-free tract might merely have been a little leg-weary, all of these young men were 'running temperatures'; the effort had been too much for them, and the concern of their old schoolmaster had been well founded. It is, indeed, well known that undue exertion may precipitate an attack in those infected with malaria. Severe endemic malaria is a drawback that even high morale cannot overcome; and, what is worse, it lowers the norm of performance.

When we turn eastwards we find ourselves again in regions of Type 1 and 6. If anywhere among the millions in Bengal, here

might have been found men of vigour; but the East India Company was looking west, not east, and since eastern Bengal, on the whole, lay off the track of armies, these plains have little military record. Perhaps for this reason, followed by that of the more recently disturbed political annals of Bengal, no effort to recruit there was seriously made to my knowledge. Lastly, in the Hill Tracts of the eastern frontier the campaigning qualities of the tribesmen have proved of service in guerrilla warfare. The difficulty now is to stop brigandage, massacre and even head-hunting, among tribesmen still armed with modern rifles and machine-guns.

Since this paper was drafted, the Census Report of India for 1941 has appeared (albeit in abbreviated form), and shortly after, the official reports upon the famine of 1943 in Bengal. These have been followed in turn by brief statements of severe crop scarcity and food shortage in many regions, many of them regions in which famine has long been a recurrent scourge. Scrutiny of the district statistics to 1941 makes it clear that had the change and variability figures been recalculated for the sixty-year period, (1) the rate of increase per cent would have been greater almost everywhere, so that the values of change to be given would have risen; while (2) the curve would have been free of any violent new rise, or certainly free of a sudden fall. The curve of each graph exemplifies these two facts by the dashed line linking the six statistical points, which is carried on to 1941. But the difference of the sixty-year calculations from those shown would only have been a matter of degree: the boundaries of the type-areas would still have stood out in much the same regional distribution.

But when the present decade comes to a close, greater differences will be apparent. Prolongation of the war, when crop failures brought disaster to Bengal, the cessation of the imports of rice from Burma and the loss of exports of jute, brought the long period of high steady increase to a catastrophic end. The area is small, but the provincial population (60 millions of India's 400 millions) is large. There were periods of scarcity elsewhere, but in none was the death roll so high. Since then, notably during 1946–7, war and failures of rain have led to crop scarcity such that the increased population now exceeds the available food supply. Inevitably, it seems, the suffering will be severe, the death roll will be heavy, the rate of change will be lowered and the variability heightened once more.

What are the psychological responses to all this cannot be said; but it seems right to ask whether the appalling 'communal' riots and strife can be adequately understood as having merely a 'religious' or 'political' significance. It is too rarely pointed out that in India the worst of the rioting has occurred either in the great modern cities, filled with rival immigrants (preponderatingly males), where the ghastly struggle of merchant-brokers, employers and employees sets the starving, the fevered and the desperate at one another's throats. The first great rural exception has been eastern Bengal, where famine and fever have taken toll among the manly Muslim peasantry who angrily begrudge the mainly Hindu landlord class their rents and the mainly Hindu 'up-country' broker class their unchecked and inordinate famine-profits. Naturally, once fanaticism is let loose, the poorest suffer with the comparatively well-to-do, and the Hindu untouchables are 'beaten up' along with their 'betters' – or rather, being unprotected, it is they who suffer most. In India the world war has left the rich richer, but the poor still poorer than before. The first budget introduced by a Government composed of Indians has brought in a new policy of severe taxation upon excess profits, coupled with a lightened burden upon a universal necessity, salt (Feb.–Mar. 1947). But the most enlightened social and economic guidance and control will be required to fill the gap between food supply and population, without which the areas of 'crisis' or 'stagnation' will inevitably spread – signs that not only the body of the Indian masses but their outlook, mind and very soul are sick.

(c) Regional applications to other lands

The great advantage of any index is of course its value for comparison. Within the variety of India's landscape, life and culture, there is throughout so much in common as to make comparison most fruitful, so long as we can find standards for comparison based on the features to be found everywhere. It is true that searching analysis demands local knowledge and intensive survey. But a basis for comparison is required before any set of analyses can be unified into an understanding of the whole; and it is still more needed in order to frame any unified policy of statesmanlike development in a region or community.

In most of the world's thickly populated lands fundamental vital experiences are statistically measured; and they can there-

fore be compared, at least from region to region, in a given country. Hence, as a geographer concerned with 'habitat' from his interest in the inhabitants, I feel that the measuring stick of 'variability', rough and ready though it may be, may prove of some service to the psychologist and sociologist in laying out the ground for their own field. For, since the *vital* experience of a society is statistically ascertainable, 'variability' gives a needed preliminary measure of profound and widespread *social* experiences. By their accumulated differences, these experiences, vital and social, give rise, in a region with a high survival rate, to confidence or natural expectancy of security, while in another region they engender the uncertainty of mind and action derived from the alternation of hopes felt in prosperity with the despair of famine, ruin or bereavement; or in a third, they lead to hopeless apathy. Hence 'variability' in population change may provide a clue to the prevalence of the states of group morale and of mind that characterise a given generation, whether of the villagers or the citizens of each of the different regions.

In particular, such an index may serve to indicate, in regions where other information is not available – or, it may be, is inadequate for effective action – the existence of catastrophe which calls for therapeutic effort. In India the areas of 'stagnation' were too long left forgotten, because they did not show the dramatic contrasts of relative well-being with misery and death which are seen in a famine crisis, when irrigation works, famine relief and charity are provided and administered. Yet 'stagnation' was, and remains, as deadly a condition as that of 'crisis', if not the deadlier of the two.

With regard to Occidental cultures, however, the descriptions and uses given and suggested for overall population indices inevitably raise the question of how far indicators of this kind are likely to be appropriate in countries where – unlike India – parenthood is not inevitably associated with sexual relations. The significance of the differences in the net reproduction rate in European cultures, for example, can hardly be said to be clear; but that they do have a deep significance in the psychological and sociological sense is widely agreed.

IV. SUMMARY

This study follows upon four previous papers: the first two on suggested geographical explanations of the index of variability in population change applied to India; a classification of Indian demographic types, and of their regional distribution and geographic causation; and finally a discussion of some special statistical considerations.

In this present paper examples of the four best-marked types of change and variability of population are given, chosen from Indian village communities known to the author; and the social and psychological significance of these regional types is outlined. To estimate the variability, a smoothed exponential curve was assumed, passing through the first and last given points available, the years 1881 and 1931, and an average deviation of the given values from this curve was determined. The effects of the 'recurrent crises' of famine and epidemic in a Deccan famine tract are contrasted with those of 'stagnation' in a tract of endemic malaria (W. Bengal); and both with the effects of half a century's 'high natural increase' (E. Bengal), and of more recent 'colonisation' in Assam and the Punjab. The broader and more certain effects, social and psychological, of these circumstances are first shown, and later, less easily verified effects, as they appear to apply to particular classes, groups and activities, are discussed.

The data and observations recorded in this paper are intended to do no more than indicate the possibilities of a technique for diagnosis and treatment of social problems in wide areas where only the crudest measures of vital phenomena exist. A subsidiary purpose has been to give some indication of the contribution which social geography might make to the basic psychological and sociological study of a region, and hence to the integration of sciences concerned with human problems.

1. Geddes, Arthur (1937) 'The Population of Bengal – Its Distribution and Changes: A Contribution to Geographical Method', *Geog. J.*, vol. 89, no. 4.
—— (1941) 'Half a Century of Population Trends in India: 1881–1931', *Geog. J.*, vol. 98, pts 5–6
—— (1942) 'Variability of Change: The Population of India as a Regional Demographic Index', *Geog. Rev.*, vol. 32, p. 4.
—— (1943) 'The Index of Variability in Population Change: Some Statistical Considerations', *Ind. J. Med. Res.*, vol. 31, p. 115.

INDIA'S POPULATION GEOGRAPHY IN AN ERA OF CHANGE[1]

A. T. A. Learmonth

INTRODUCTION

In India's 1961 census, almost a million enumerators recorded a population of about 439 millions. The rate of increase 1951–61 was estimated at about 2·2 per cent per annum – well in excess of the expected 1·3 per cent, and hence largely accounting for smaller gains in per capita income than had been hoped for under the second five-year plan. No doubt there has been something of a demographic revolution, though antecedents can be traced in many parts of the country as far back as 1921 or even farther. At the time of writing this paper, towards the end of the period between the 1961 and 1971 censuses, it is necessary to be rather tentative in reviewing current trends.

The writer believes, though by no means all authorities agree, that a major factor in the change has been successful campaigns for malaria control and at present towards malaria eradication, which have been associated with increases in birth rates – malaria, like some other infections, tends to increase abortion rates – as well as decreases in death rates. The issue is controversial. Ceylon, where malaria control and the demographic revolution were established early, has often been used to suggest a clear relationship between malaria and natality; but recent retrospective surveys point out puzzling discrepancies in the dates of malaria control *vis-à-vis* the population upsurge. Malariologists are somewhat sensitive when they are – as it were – blamed for the population explosion, and point out that there have been many concurrent social and economic changes as well as the campaigns against the malaria vectors. In places, at least, the control of malaria has allowed abandoned land to be reclaimed, or fresh land won for tillage – the fruit of the looked-for rise in economic activity once the economic drag of malaria is removed, which is indeed a reality in rural areas if there are resources of land available to bring into cultivation. Clearly such gains must be fed into the balance sheet. And the figures must be viewed in perspective.

India's rate of population increase is by no means the highest in the world, though it applies to a formidably large total population, in fact about a sixth of the world's people. Her rate of increase has at times been equalled or exceeded by countries now industrialised and with high standards of material living. Yet by common consent many parts of India are overpopulated – difficult as the concept is to quantify – and not only in rural areas but very significantly in urban areas also. (Urban areas, for instance, have high rates of unemployment and underemployment – 17 per cent and 16 per cent respectively, whereas an estimate for rural areas is 7 per cent and 15 per cent – though criteria may be sharper in towns to the people concerned.)

Campaigns for family planning and birth control have government support. They seem timorous compared with the massive assemblage of propaganda, visual aids and contraceptive devices made available in China, according to reliable Indian observers, at least when population control was compatible with the current party line in China some years ago. But early experiments in urban and rural areas were not wholly successful; the rhythm method, everywhere suspect, could not be applied by illiterate people, and the costs of contraceptives have proved prohibitive to lower-income groups even in towns. Encouragement of male sterilisation, through bounties and propaganda, has had limited success. Of the intra-uterine contraceptive devices, the plastic loop can be made cheaply in India and may effect a change in the demographic prospects; or a cheap and effective oral contraceptive may appear. But the very acceptance of the idea of family limitation involves complex problems of mores – something of Mr Doolittle's middle-class morality. There is some evidence that in the present middle classes smaller family size was associated with a lower income limit of about £300 per annum over a decade ago; allowing for changes in the value of money, incomes of that order may well be widespread if the five-year plans are successful. The socio-economic groups reaching this modestly prosperous income level in the coming decade may not, or may not immediately, react in family limitation as do the present middle classes. There may well be at least a lag, and perhaps a long lag. So far evidence for lower fertility rates in towns is surprisingly slender, perhaps because of the greater ease of access to health services. On a general view, then, a combination of successful economic development leading to higher family incomes, and

widely available, acceptable and understood contraceptives, appears to offer the best prospect of a slowing down in population growth which will facilitate further advances in standards of living. Some regional analyses, if only by example, may add understanding and insight to these generalisations.

In 1941 and 1942, Geddes published analyses of regionally different types of demographic variations in the Indian subcontinent, work which may now be regarded as classic. The original papers are still essential basic reading, and an extension of his work into the sociological and psychological implications is produced as the preceding paper. This interdisciplinary projection of his work lay very near to Geddes's heart and was often uppermost in his mind. At the stage when he wrote, the work was necessarily descriptive, as he well knew; great efforts were made to get support for an interdisciplinary project which would follow up this suggestive paper more analytically, but the necessary funds could not be obtained. Towards the end of his life Geddes indignantly rejected a phrase in the present author's original version of this paper, published in Australia, suggesting that his sociological and psychological hypotheses are necessarily at hazard for lack of evidence. For him the facts of high infant mortality, frequent bereavement, high orphan rates and the like were so stark and quantitatively established as to be unquestionably of sociological and psychological importance; thus for him, the paper is by no means speculative or unproven, even though the effects on the society and psyche have only been assessed qualitatively rather than measured by the appropriate disciplines. Through the demographic revolution discussed later in the paper, Geddes's regionals may remain of socio-psychological and even of socio-economic significance.

Fig. 33.8 is a classic synthetic map of malaria before the eradication campaigns of the last two decades, based on work by Christophers and Sinton; this should be compared with Geddes's map of demographic regions. Even purely cartographic correlation between the two maps justifies Geddes's emphasis on malaria, particularly hyperendemic malaria, as a major factor in his population-stagnation type of graph and demographic region: his regions of recurrent crisis are largely areas of epidemic malaria, drought and famine, but epidemics of plague and influenza have also been important.

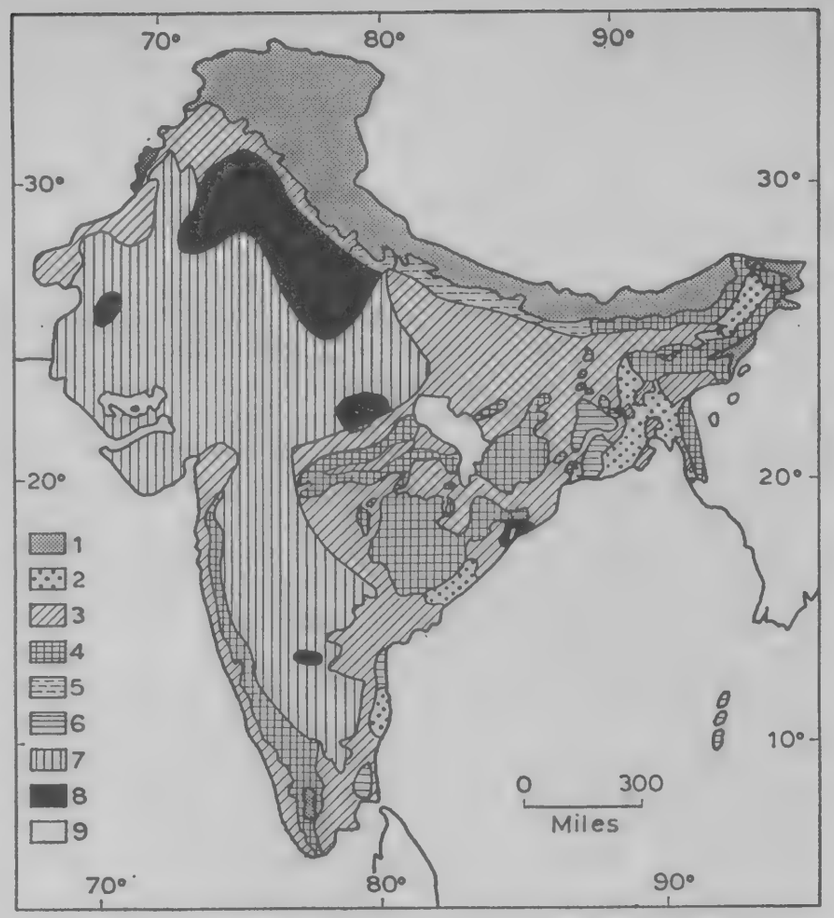


Fig. 33.8 Malaria regions, 1938. The heavy line marks the broad division between endemic and epidemic areas. (1) Areas over 5000 ft (non-malarious); (2) known healthy plans (spleen rate under 10 per cent); (3) more or less static moderate to high endemicity, intensity depending on local factors – seasonal variations moderate, fulminant epidemics unknown; (4) hyperendemic jungly hilly tracts and *terai*; (5) probably hyperendemic hill areas; (6) hyperendemic other than hills; (7) variable endemicity of drier areas, usually with autumnal rise in fever incidence (potential epidemic areas), spleen rate low except in years following epidemics or in special local circumstances, much affected by irrigation conditions; (8) known areas liable to fulminant epidemicity (diluvial) malaria, spleen rate high during and immediately after epidemics, slowly falling to low rates in about five years; (9) unsurveyed

Source: *Annual Report of the Public Health Commissioners for 1940* (Delhi, 1941), based on work by S. R. Christophers and J. A. Sinton

In the ensuing discussion, Fig. 33.9, a map of population increase 1951–61, has been used to complement Fig. 33.8 of demographic

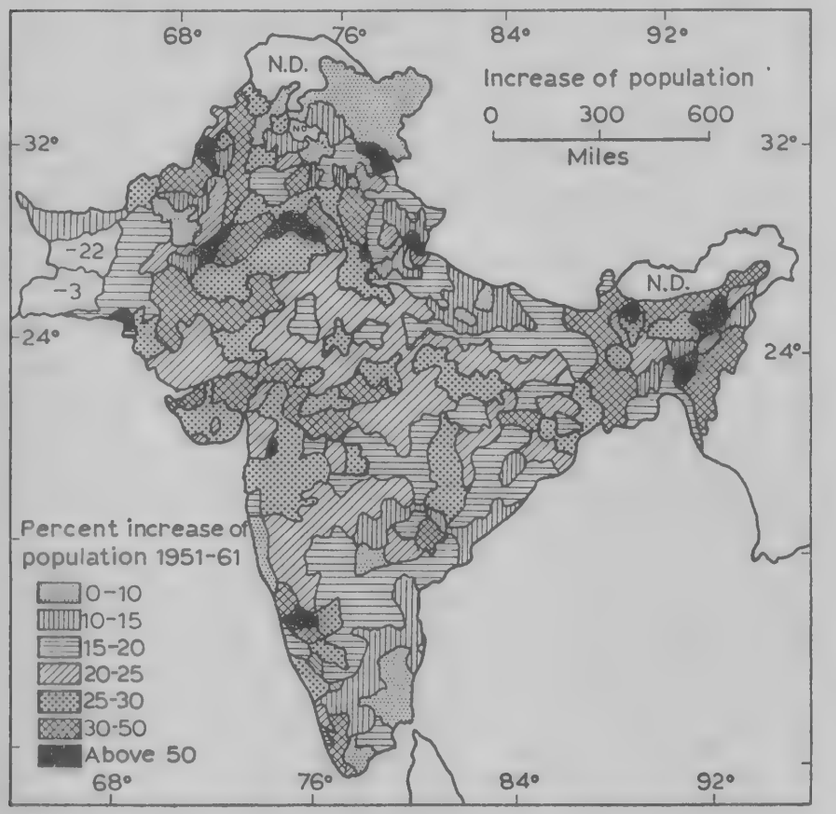


Fig. 33.9 Increase of population, 1951–61
Source: Census of India, 1961; Census of Pakistan, 1961

regions 1881–1931; we may note parenthetically that a somewhat different type of cartographic analysis for the period 1921–41 showed that already areas of population decrease were much fewer, and that some of Geddes's regions of population stagnation were being eroded, apparently by urban/peri-urban and industrial development, e.g. in the Hooghlyside belt north and south of Calcutta–Howrah. An analogous complement to Fig. 33.8, of malaria regions, need not be provided: one need only picture a rather rapid spread of malaria control from a tract round Dharwar to cover virtually the whole of India, and replaced by eradication campaigns from 1957 onwards. It will be assumed that the reader has access to a map showing the high population densities, often over 1000 per square mile – even as rural densities – and going up to 2000 and even more locally, in parts of the Bengal delta and the west coast, especially the ricelands of Kerala. Fig. 33.9 shows that by 1951–61 no administrative district maintained population decrease. (Interestingly enough, however, quite a number of small towns showed marked decreases; in this respect India shows consonance with trends in advanced countries with well-developed transport systems. In India bus and lorry services are of considerable importance for short-distance and even some long-distance traffic, and even now they may be effecting changes in the range of traffic to major markets, causing decay in smaller market and service centres.)

Geddes's regions of stagnation were mainly in the ricelands of the western 'dying' delta of Bengal, the lower and middle Ganga plains, parts of the east coast plains including the Kaveri delta and parts of the malarious west coast and Ghats edge. In places there is now very high increase, by high industrial and refugee immigration to Hooghlyside, probably more by natural increase in north Bengal, and certainly in western Mysore. Yet the earlier population stagnation still leaves residuals, in continuous tracts of *relatively* low population increase (well below the national average), in the Ganga plains and in a large part of south-eastern India. This last was partly in the large south Indian stretch of high natural increase and rather comparable 'variable increase', in 1881–1931, largely by reason of its gradual recovery from the great famine of 1878, just before the first fairly reliable census; much of this is tank irrigation country, with small to medium reservoirs the best of which extend the watering season (e.g. for rice), and cope with minor irregularities in the rains, but do not protect the farmers against major failures of rainfall or permit the storage of water from a good to a poor rainfall year. Much of this tract has been afflicted by fitful rainy seasons in recent years, causing famine or near-famine conditions in the years immediately following independence, and again at the time of writing this paper. Tamil- and Telugu-speaking areas in particular (as against Kannada or Malayalam areas) have long had a tradition of emigration, particularly among economically vulnerable low-caste groups, and this increases after each drought year and particularly after a

series of drought years, often increasing the urban unemployment/underemployment referred to earlier, particularly in the million cities of south India – Madras, Bangalore and Hyderabad.

It is remarkable that the very densely peopled rice-lowlands and tapioca on laterite upper plains of Kerala supported 'high natural increase' 1881–1931, and continue to record large population increases. The introduction of tapioca on a large scale in 1919 increased the subsistence basis, and this is an area of considerable diversity mainly of small and handicraft industry. It is one of the clearest cases of overpopulation, none the less poignant for the good education and self-respect of the people. The large proportions of literacy, of newspaper-reading, of Christianity, of caste-division and of communism are remarkable. (After the 1957 elections, when the communists gained ascendency in the states legislature, it was jokingly rumoured that a new caste was gaining in power – 'Roman Catholic Party Member'.)

Geddes's recurrent crisis region covers mainly the variable rainfall tracts of the interior of the Deccan and northern Gujarat, and Rajasthan including the Thar desert. A sort of aureole around this tract includes much of the intermediate type. It is striking how much of this area is covered by increases above the national average in 1951–61, and patchily with very high increases. Few of the high increases are associated with irrigation works (e.g. the Rajasthan canal from the Sutlej, now partly functioning), or with industry (perhaps round Ahmedabad); in the south of the Punjab and north of Rajasthan, refugee immigration may be significant. Near-famine conditions prevailed locally in 1964–5, but in the period surveyed the suggestions are of surprisingly uniform increases from a starting point of patchy conditions of epidemic, drought and near-famine, variable in both space and time.

Geddes's areas of 'colonisation' in the Punjab – now partly in West Pakistan and partly in India – still show some very high increases; the complementary area of high natural increase 1881–1931 along the long intensively cultivated well-irrigation area along the hillfoot shows more sporadic increases now. The considerable area of high natural increase in relatively sparsely peopled country in the outer Himalayas north-east of Delhi, doubtless financed as it were largely by the potato (as in nineteenth-century Ireland), now shows relatively low increases, perhaps because of emigration.

The remarkable area of 'colonisation' in the Assam valley 1881–1931 was largely of family-by-family movement of Muslims from the 'high natural increase' area in the active delta of east Bengal, now East Pakistan – where sheet-floods cause damage to property and crops but renew soil fertility and do not suit the locally predominant malaria vector. Despite the virtual cessation of this movement after partition, and probably even some return movement of refugees from India to East Pakistan, the Assam valley continues to show very high rates of population increase, probably due both to immigration to farms and tea-estates and to an upsurge in natural increase due to malaria control (as also in the surrounding hill tracts).

Thus the well-marked demographic regions mapped by Geddes have been succeeded by a much more blurred picture in which there are only residual areas of relative low (but still generally considerable) population increase. Though the writer believes that malaria control and eradication represent the greatest single factor in this demographic revolution, there are other strands of importance in the complex and constantly changing tapestry.

It is a matter of common observation that there is a large movement of population towards the cities and particularly large cities, still tending towards predominance of young men in the north, causing many consequential social and health problems – poor slum and shanty-town dwelling, prostitution, the spreading of tuberculosis and venereal diseases in urban slums and back into rural villages. Though there is more family movement in the south, it is still often to poor mud and kerosene-can shanties, huddled in ill-drained and insanitary hollows among more prosperous dwellings. Despite the fearsome toll of suffering from urban immigration out of step with economic developments – impressive as these are in places – many observers in India and abroad hope for more and yet more urban-industrial development as the only hope for ultimately better standards of living and a well-founded population stability to replace the Malthusian checks of Geddes's stagnation type of demographic curve.

Community development programmes have attained considerable successes – more easily, however, in amenities than in productivity; locally too, land reform has overcome various legalistic evasions sufficiently to benefit many former tenants, now holding directly from the state, and to increase productivity. The winning

of new land or reclamation of abandoned land does go on and locally affords a useful safety valve for population increase or for refugee immigrants – in the formerly malarious terrains of the Nepal *terai*, the forested hilly tract or Malnad east of the Western Ghats, the Dandakaranya project area of the Bastar highlands north-east of the Godavari delta, the Assam valley and a few other areas. Though locally significant, numbers involved are rather small, and greater importance seems to lie with more intensive methods in existing agricultural tracts, and with reduction in the depredations of rodents and other vermin. On the other hand rural population persistently increases so that holdings per capita diminish, and measures of intensification are only locally strikingly successful. Such measures of standards of living and diet as are available indicate rather better conditions in the north, roughly in Punjab, the new state of Hariana, and western Uttar Pradesh (though the sharper chills of the cool season demand more food and clothing there, and better housing), with a gradient to poorer standards of living in the east and especially the south, where the economically vulnerable groups are extremely liable to fall below the famine line in drought years.

It now appears that the very high urban increase of 1941–51 was largely attributable to the great population transfers which accompanied and followed the tragic carnage of partition, when almost 20 million people moved – about equal numbers from each of the new countries – and at least 750,000 were killed, in communal rioting. Urban population increased: 1941–51–61, 43·81 million to 61·87 million (41·2 per cent), to 79·24 million (28·1 per cent – only a little more than the general population increase). The Registrar-General points out that after allowing for the redefining of towns in 1961 the urban increase was 21 millions or 36 per cent; this is twice the rural rate but still, he thinks, much less than the impression or expectation of urban explosion which had been built up during the decade. Given the high rates of urban unemployment already quoted, the urban-industrial increases are not yet solving the vexatious problems of the countryside, and this despite the great changes which must come to many readers' minds: the stimulus of hard-working and energetic refugees in Ludhiana with its new bicycle and sewing-machine plant, and the little workshops by the roadsides making car spares and complementing the older woollen industries of the town; the Sindri fertiliser plant, the power stations and steel works of the Damodar valley coalfield, and the T.V.A.-type developments in its hinterland; the new steel towns of Durgapur, Rourkela and Bhilai and the expansion and diversification in the now fifty-five-year-old steel town of Jamshedpur-Tatanagar, named from its visionary Parsi founder's family. Or again the shipyards at Visakhapatnam, the heavy electricals plants at Bhopal, or the group of new central government factories at Bangalore (aircraft, aero-engines, machine tools, electronics, telephones), complementing the earlier state government and private-sector factories and in turn gathering round them the beginnings at least of a congeries of ancillary small factories and workshops. Or the continued activity and growth in the great complex foci of earlier private, partly foreign-owned, industry in Bombay, the Ahmedabad area, Hooghlyside, Madras, Kanpur, and to a less extent in most – though not quite all – of the more isolated 'up-country' centres. Fig. 33.10 summarises these areas where we may look for hopeful socio-economic and demographic change, more obvious so far, as we have already noted, in improved amount and stability of income from the level of factory worker upwards than in reduced family size; and so far with no solution to the problems of the slum-dwelling casual daily labourer or coolie or mazdoor – the scavenger, the porter, the truly menial worker.

In total these industrial activities would represent a considerable source of wealth – to a population the size of Australia's – and yet are increasing far too slowly to give satisfactory rises in standards of living for India's 500 million people. Yet there are now large if not ample foundations of experience and skill; the growing numbers and sizes of medium to large towns and cities – though still mostly weak in municipal spirit, funds and authority, the still prosperous railways, busy and well-run internal and external airlines, the growing passenger and goods services by road, the growing coastal and overseas shipping, all equally are foundations with which India is better endowed than many underdeveloped countries. There are serious drags – the disputes with China and Pakistan, the tortuous and still caste- and faction-ridden politics even – or especially, within the ruling Congress party, the Parkinsonian profusion of the clogs to administrative efficiency at all levels, quickly reproduced even in new bodies like the planning commission. But there is little doubt that, purely technically, economic

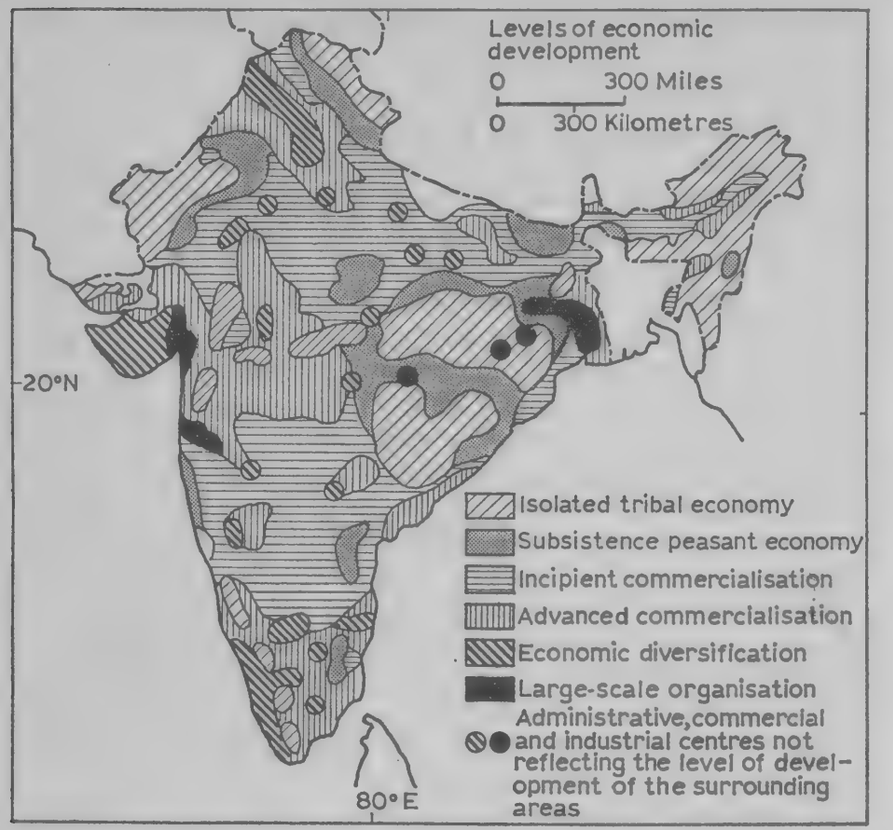


Fig. 33.10 Levels of economic development

Source: J. E. Schwartzberg, *Annals of the Association of American Geographers*, vol. 62 (1962). Schwartzberg illustrates three methods of mapping economic development – in this map using subjective integration

development could feasibly provide sufficient increases in the standard of living, accompanied by widespread literacy and higher education, on a scale sufficient to justify hopes of a slowing down in population growth rates. Unless there is in population control as dramatic a *deus ex machina* as the residual insecticides have proved for malaria control in this subcontinent – though not all over the world – we may expect population stability to lag well behind the economic developments and the improved standards of living, perhaps by a generation. So we may, in all probability, look forward to three very difficult decades for India's people – and therefore for the world at large, since we are considering a sixth of the population of the globe – between now and the end of the twentieth century.

Note

1. The substance of this paper was published in the *Australian Journal of Politics and History*, vol. 12, no. 2 (August 1966) 146–54, under the title 'Selected Aspects of India's Population Geography'.

Although rapid population growth is now seen primarily as a 'problem' in non-Western societies, two hundred years ago the West too was concerned with it. One of the most famous commentators was an English parson, Thomas Malthus (later professor at the East India College at Haileybury), who in 1798 published an 'Essay on the Principle of Population'. The essay won immediate acclaim and was a focus of controversy for the next thirty or forty years. Malthus's basic position is contained in the extract printed here. It was a pessimistic one, reminiscent of much that is being said at the present day. How far it is applicable to contemporary developments is taken up by Hla Myint in a chapter reprinted below from his book 'The Economics of Developing Countries'.

A crucial question in a study of the population explosion is whether some optimal population size exists in different countries in the sense of enabling them to achieve their economic objectives most efficiently. The most pressing objective of the developing countries is to raise their rate of economic growth and development, and population size and growth is an important factor in this process. For this reason we have included Professor Arthur Lewis's discussion of the problem from his 'Theory of Economic Growth' under the title 'What is the Right Size of Population in Relation to Resources?'

ON POPULATION AND SUBSISTENCE

Thomas Malthus

(from *An Essay on the Principle of Population* (1798))

THE principal object of the present essay is to examine the effects of one great cause intimately united with the very nature of man. . . . The cause to which I allude is the constant tendency in animated life to increase beyond the nourishment prepared for it.

It is observed by Dr Franklin that there is no bound to the prolific nature of plants or animals but what is made by their crowding and interfering with each other's means of subsistence. . . . Through the animal and vegetable kingdoms Nature has scattered the seeds of life abroad with the most profuse and liberal hand; but has been comparatively sparing in the room and the nourishment necessary to rear them. Plants and irrational animals . . . are all impelled by a powerful instinct to the increase of their species; and this instinct is interrupted by no doubts about providing for their offspring. Wherever therefore there is liberty, the power of increase is exerted; and the superabundant effects are repressed afterwards by want of room and nourishment.

The effects of this check on man are more complicated. Impelled to the increase of his species by an equally powerful instinct, reason interrupts his career, and asks him whether he may not bring beings into the world for whom he cannot provide the means of support. If he attend to this natural suggestion, the restriction too frequently produces vice. If he hear it not, the human race will be constantly endeavouring to increase beyond the means of subsistence. But as, by that law of our nature which makes food necessary to the life of man, population can never actually increase beyond the lowest nourishment capable of supporting it, a strong check on population, from the difficulty of acquiring food, must be constantly in operation. This difficulty must fall somewhere, and must necessarily be severely felt in some or other of the various forms of misery, or the fear of misery, by a large portion of mankind.

That population has this constant tendency to increase beyond the means of subsistence, and that it is kept to its necessary level by these causes, will sufficiently appear from a review of the different states of society in which man has existed. But, before we proceed to this review, the subject will, perhaps, be seen in a clearer light if we endeavour to ascertain what would be the natural increase of population if left to exert itself with perfect freedom; and what might be expected to be the rate of increase in the productions of the earth under the most favourable circumstances of human industry.

In the northern states of America, where the means of subsistence have been more ample, the manners of the people more pure, and the checks to early marriage fewer than in any of the modern

states of Europe, the population has been found to double itself, for above a century and a half successively, in less than twenty-five years. . . .[1] In the back settlements, where the sole employment is agriculture, and vicious customs and unwholesome occupations are little known, the population has been found to double itself in fifteen years. Even this extraordinary rate of increase is probably short of the utmost power of population. . . . According to a table of Euler, calculated on a mortality of 1 in 36, if the births be to the deaths in the proportion of 3 to 1, the period of doubling will be only 12 years and 4–5ths. . . . Sir William Petty supposes a doubling possible in so short a time as ten years.

But, to be perfectly sure that we are far within the truth, we will take the slowest of these rates of increase, a rate in which all concurring testimonies agree, and which has been repeatedly ascertained to be from procreation only.

It may safely be pronounced, therefore, that population, when unchecked, goes on doubling itself every twenty-five years, or increases in a geometrical ratio.

The rate according to which the productions of the earth may be supposed to increase, it will not be so easy to determine. Of this, however, we may be perfectly certain, that the ratio of their increase in a limited territory must be of a totally different nature from the ratio of the increase of population. A thousand millions are just as easily doubled every twenty-five years by the power of population as a thousand. But the food to support the increase from the greater number will by no means be obtained with the same facility. Man is necessarily confined in room. When acre has been added to acre until all the fertile land is occupied, the yearly increase of food must depend upon the melioration of the land already in possession. This is a fund, which, from the nature of all soils, instead of increasing, must be gradually diminishing. But population, could it be supplied with food, would go on with unexhausted vigour; and the increase of one period would furnish the power of a greater increase the next, and this without any limit. . . .

Let us consider at what rate the produce of this island might be supposed to increase under circumstances the most favourable to improvement. If it be allowed that by the best possible policy, and great encouragements to agriculture, the average produce of the island could be doubled in the first twenty-five years, it will be allowing, probably, a greater increase than could with reason be expected. In the next twenty-five years, it is impossible to suppose that the produce could be quadrupled. It would be contrary to all our knowledge of the properties of land. The improvement of the barren parts would be a work of time and labour, and it must be evident to those who have the slightest acquaintance with agricultural subjects that, in proportion as cultivation extended, the additions that could yearly be made to the former average produce must be gradually and regularly diminishing. . . .

It may be fairly pronounced, therefore, that, considering the present average state of the earth, the means of subsistence, under circumstances the most favourable to human industry, could not possibly be made to increase faster than in an arithmetical ratio.

The necessary effects of these two rates of increase, when brought together, will be very striking. Let us call the population of this island eleven millions; and suppose the present produce equal to the easy support of such a number. In the first twenty-five years the population would be twenty-two millions, and the food being also doubled, the means of subsistence would be equal to this increase. In the next twenty-five years the population would be forty-four millions, and the means of subsistence only equal to the support of thirty-three millions. In the next period the population would be eighty-eight millions, and the means of subsistence just equal to the support of half that number. And, at the conclusion of the first century, the population would be a hundred and seventy-six millions, and the means of subsistence only equal to the support of fifty-five millions, leaving a population of a hundred and twenty-one millions totally unprovided for.

Taking the whole earth, instead of this island, emigration would of course be excluded; and, supposing the present population equal to a thousand millions, the human species would increase as the numbers 1, 2, 4, 8, 16, 32, 64, 128, 256 and subsistence as 1, 2, 3, 4, 5, 6, 7, 8, 9. In two centuries the population would be to the means of subsistence as 256 to 9; in three centuries as 4096 to 13, and in two thousand years the difference would be almost incalculable.

In this supposition no limits whatever are placed to the produce of the earth. It may increase for ever and be greater than any assignable quantity; yet still the power of population being in every period so much superior, the increase in the human species

can only be kept down to the level of the means of subsistence by the constant operation of the strong law of necessity, acting as a check upon the greater power.

Malthus goes on to summarise his theory in three propositions, as follows:

1. Population is necessarily limited by the means of subsistence.
2. Population invariably increases where the means of subsistence increase, unless prevented by some very powerful and obvious checks.
3. These checks, and the checks which repress the superior power of population, and keep its effects on a level with the means of subsistence, are all resolvable into moral restraint, vice and misery.

Note

1. The reader must bear in mind that this was written in 1798.

POPULATION PRESSURE IN UNDERDEVELOPED COUNTRIES

H. Myint

An underdeveloped country in its traditional state of isolation tends to have a fairly stable population with a high birth rate and a high death rate balancing out over the long run. Although accurate information is not available, the most probable estimate seems to be that, before their contact with the West, many underdeveloped countries in Asia and Africa had high birth rates in the region of 4 per cent per annum matched by equally high death rates of about 4 per cent. Starting from the initial position of high birth and death rates, I shall describe the typical pattern of population growth with three stages or sets of factors that reduce the death rate while the birth rate continues at its high initial level.[1]

First, the introduction of a modern administration improving law and order and eliminating local warfare, and the introduction of modern transport and communication eliminating local famines and spreading trading in foodstuffs, can by themselves reduce the death rate of an underdeveloped country by as much as 1 per cent per annum. So if the birth rate remains unchanged, population will be increasing at 1 per cent per annum. Secondly, the introduction of public health measures controlling epidemic and endemic diseases such as plague, smallpox, cholera, malaria, yellow fever (and eventually tuberculosis) can reduce the death rate by at least another 1 per cent. So if the birth rate remains the same, the population of the underdeveloped country will be increasing at some 2 per cent per annum at the end of this second stage. Thirdly, the spread of individual medical attention and the increase in the number of doctors and hospitals can bring down the death rate by another 1 per cent, depending on the age structure of the population. In this final stage, therefore, if the birth rate remains the same at 4 per cent and the death rate is reduced to 1 per cent per annum, the maximum possible rate of growth in population will approach 3 per cent per annum.

Most of the underdeveloped countries have passed through the first stage and are either towards the end of the second stage with the rate of population increase between 1½ per cent and 2 per cent per annum or are well into the third stage with the rate of population increase over 2 per cent. The fastest rate of population growth is to be found in Latin America with an average annual rate of growth of 2·4 per cent, while Asia and Africa are fast approaching the 2 per cent rate of growth. India, which until the last census in 1961 had a population growth of 1·4 per cent, now seems to have passed the 2 per cent mark, mainly due to better control of malaria. There are still a few Asian countries at the lower end of the 1½–2 per cent range such as Burma, with her high rate of infant mortality, and possibly China which many authorities believe to have a relatively slow rate of population growth. But for most of the underdeveloped countries, a 2 per cent rate of population growth is typical, giving us a possibility of the population doubling every thirty-five years.[2] So when people speak of the 'population explosion' in the underdeveloped countries, they mean that the populations of these countries are not only increasing but increasing at a faster rate of growth year by year with a high rate of acceleration. An extreme example of this is the case of Ceylon where D.D.T. has wiped out malaria, reducing the death rate from 2·2 per cent to 1·2 per cent in seven years from 1945 to 1952 – a fall which took seventy years in England and Wales.[3]

Although the rate of population growth is clearly very impor-

tant, two other factors must be taken into account in studying the problem of population pressure in the underdeveloped countries.

The first is the population density in relation to natural resources and technology. Population density in a significant sense cannot be measured simply by taking the number of people per square mile. It will depend on a number of factors such as the quality of agricultural land, climate and water supply, the existence of mineral resources, the possibilities of hydro-electric power, to name a few; and also, very importantly, on the level of productivity and technology. Europe and Asia, for example (excluding U.S.S.R. from both), are estimated to have the same amount of agricultural area per person of 1·5 acres, but with widely differing standards of living or 'population densities' in the economic sense. But taking the level of technology to mean the indigenous methods of agricultural production and making allowances for obvious differences in natural resources, it is still possible to say that there are very significant differences in population densities between the different parts of the underdeveloped world. Compared, for example, with the agricultural area per person of 1·5 acres in Asia, Latin America is estimated to have 6·9 acres, and Africa 10·6 acres of agricultural area per person.[4] In a relatively sparsely populated region like Latin America, even a high population growth rate will not for the time being give rise to the problem of population pressure. In fact, many would regard it as a favourable factor for future economic development. In contrast, with the dense population of China even a small population growth rate will create formidable problems of population pressure.

There are two main types of underdeveloped country with high population densities. Firstly, there are the long-settled agricultural countries such as India, China or Egypt. In a sense, their present population problem is due to their successful development as agricultural countries in the past. This enabled them to support a dense population before it started growing in the modern period. Secondly, there are the island plantation economies, such as the West Indies, Fiji, Mauritius or Ceylon. Here, immigrant labour has been brought from outside to work in the mines and plantations to add to the indigenous population. Population pressure subsequently developed, helped by the relative ease with which diseases such as malaria can be wiped out from the islands. In general, the first group of countries tend to have a larger population base than the second – contrast India's 438 millions in 1961 with Mauritius's 0·6 million.

This brings us to the second important factor to be taken into account in studying the population problems of the underdeveloped countries, that is the absolute size of the population base. When giving aid, countries like India with a huge population base clearly present a more formidable problem than countries with a smaller population, even if the smaller populations are increasing at a faster rate. Even before 1951, when the rate of increase was relatively slow, India's population was increasing at the rate of 5 millions a year; with the new, higher rate of 2·15 per cent per annum shown by the 1961 Census, it will be increasing at the rate of 8 millions a year. The size of the population base, however, is also of great importance as it affects the total scale of the economy; this is relevant for the success of industrialisation programmes. Of course, the size of the domestic market of a country does not depend only on the numbers, but also on the income level of its inhabitants. But given the same low level of per capita incomes, a country like India offers a more favourable environment for setting up heavy capital-goods industries which depend so much on the economies of scale for their success. In contrast, a thickly populated country with a small population base such as Ceylon seems to be especially handicapped by the smallness of its market.[4] Finally, considering the absolute size of the population base serves to remind us that while the large populations of the advanced countries have grown up after, and as a consequence of, economic development, the large populations of the underdeveloped countries exist before development: this makes development not only more desirable but also more difficult.[5]

In broad terms, the population problem of the densely populated underdeveloped countries is a vivid illustration of the Malthusian theory. On a closer analysis, I would add three qualifications.

(i) The Malthusian proposition that the growth of population will exceed the growth of food supplies underrates technological innovations in food production, including the opening up of new territories and international trade. That was why the advanced Western countries were able to escape from its direful predictions. In a sense, the theories of economic development for the overpopulated underdeveloped countries are built on this

qualification. They are based on the idea that while there is a maximum limit to population growth, say 3 per cent per annum, the possible rate of increase in output owing to technical innovation and capital accumulation can exceed this maximum rate of population growth, keeping the Malthusian devil at bay until the birth rates in the underdeveloped countries begin to fall, as they have fallen in the European countries and are falling in countries like Japan.

(ii) The Malthusian theory is based on too rigid a causal connection between the rate of growth in food supply and population growth. According to the theory, population will not grow unless the food supply or the standard of living rises above the minimum subsistence level. But as we have seen, the population in many underdeveloped countries has grown very rapidly without any appreciable rise in per capita income, mainly owing to medical improvements reducing death rates. If the Malthusian theory has underrated the possibilities of technical improvements in food production, it has also underrated the medical improvements reducing death rates. The possibility that even a population struggling at a minimum subsistence level could grow in the short run simply by medical improvements entirely unconnected with the general standard of life adds to the seriousness of the problem in the modern setting. There is a gleam of hope, however, if we are willing to believe that the technical improvements in medicine which are at present reducing death rates only can be eventually extended to reduce birth rates by the invention of simpler and cheaper methods of birth-control.

(iii) The Malthusian theory is formulated in the setting of a more or less developed wage economy. According to it, with given natural resources and technology, the saturation point in a country's population will be reached when the level of wages, determined by the marginal product of labour, is equal to the minimum subsistence level. But in many underdeveloped countries the problem of population pressure takes place in the setting of the subsistence economy before the development of the money and the wage economy. In this setting, with an extended family system, many people whose marginal product on the land is below the subsistence level can continue to exist at that level because the total output from land is shared among the members of the extended family. That is to say, with the same natural resources and technology, the saturation point of population in a subsistence economy is larger than that of a wage economy (without public poor relief) because it is determined by the equality of the average product of labour to the minimum subsistence level. Many people who would have been unemployed and starved in a wage economy are maintained in 'disguised' unemployment by their relatives as part of a traditional social security system. As we shall see, much has been made of this phenomenon of 'disguised unemployment' in the post-war theories of economic development as a concealed source of saving. But there is one point which is not always appreciated. In so far as we believe that economic development can take place only through the introduction of a modern money and wage economy which entails the dismantling of the traditional social security system of the subsistence economy, the overpopulated underdeveloped countries in a process of transition have to meet a double problem of population increase. In addition to their natural rate of increase in population they also have to find jobs for those who are displaced from the subsistence economy when the development of the exchange economy and economic individualism reduces the size of the 'family' as a social security unit.

To sum up, the problem of population pressure in the underdeveloped countries has three aspects: (i) the rate of growth of the population; (ii) the existing density of population; and (iii) the overall magnitude of the population base. In these terms, the population situation in the underdeveloped countries shows considerable variation. At one end of the scale there are the underdeveloped countries which are relatively thinly populated and are not yet suffering from population pressure. These include some of the South-East Asian countries such as Burma, Thailand, Malaysia and the Outer Islands of Indonesia; many of the African countries with the notable exception of Egypt; and most of the Latin American countries. The magnitude of the population base in these countries is also small and is below 50 millions, with the exception of Brazil, which is, however, still thinly populated. At the other end of the scale there are the densely populated countries suffering from varying degrees of population pressure. These include islands with plantation economies such as the West Indies, Mauritius and Ceylon where the population pressure is acute although the magnitude of the population base is fairly small; and

the thickly populated countries of Asia and the Middle East with much larger population bases. The population pressure problem is most formidable in countries like India and Pakistan which are not only densely populated, but have a large population base and are now entering into a phase of rapid population growth above the 2 per cent annual rate.

Post-war theories about the developing countries tend to be dominated by the model of the overpopulated countries. But large areas of the underdeveloped world such as Latin America, South-East Asia and various parts of Africa started with sparse populations in relation to their natural resources and are not yet suffering from population pressure. We need, therefore, at least two main theoretical models, one for the countries with no significant population pressure, and another for the overpopulated countries. For certain purposes involving the analysis of the economies of scale, we shall find it useful to subdivide the overpopulated countries into two groups: those like India with a large population base, say over 50 millions; and those like Ceylon and Mauritius with a small population base.

Notes

1. Cf. W. A. Lewis, *The Theory of Economic Growth* (London, 1955) chap. 6, part of which is the final extract in this chapter.
2. United Nations, *Demographic Yearbooks* (Geneva).
3. P.E.P., *World Population and Resources* (London, 1955).
4. J. R. Hicks, *Essays in World Economics* (Oxford, 1959).
5. S. Kuznets, 'Underdeveloped Countries and the Pre-Industrial Phase in the Advanced Countries', *World Population Conference* (Rome, 1954) vol. 5 (United Nations, New York, 1955).

WHAT IS THE RIGHT SIZE OF POPULATION IN RELATION TO RESOURCES?

W. A. Lewis

THIS is not primarily an economic question. One may ask, for example, what size of population would maximise output per head. There is not much hope of getting a precise answer to this question because of the great number of *cetera* upon which output depends; but the question is valid and meaningful. It cannot be assumed, however, that the right population is that which would maximise output per head. A country may prefer to have a smaller population than this; because of the alleged advantages of being a small nation, the compactness of the people, easier national unity and freedom from external political responsibilities; or because it dislikes the processes by which population is enlarged – immigration, or the having of large numbers of children. Or alternatively it may be desired to have a population larger than that which maximises output per head; because of its value in defence or in aggression; or because of a desire to play a big part in world affairs; or because of a desire to accommodate immigrants, especially if they are refugees from religious or political persecution; or because of a liking for the company of other people in general, or of large numbers of children in particular. Thus the question what is the right size of population raises issues well beyond the scope of economic analysis.

Confining ourselves to economic aspects we observe four different senses in which the term 'overpopulation' is sometimes used. First, a country is said to be overpopulated if it would have a larger output per head with a smaller population. Secondly, the term sometimes means no more than that the population is larger than can be fed without importing food. Thirdly, the term is used in an extreme sense, indicating that the country's population is so large relatively to its resources that a change in population would have no effect on total output. Finally, the term is used in a vague sense to indicate that a country is using up irreplaceable resources at an excessive rate. Let us dispose of this last sense first, since, as we shall see, it leads to no positive conclusions.

Other things being equal, the size of the population determines the rate at which mineral resources are used up. The more we consume of oil, or coal, or iron, or tin, or other minerals, the less there is left in the earth's surface for future use. Can we establish a 'right' rate at which these resources should be consumed?

Resource conservation raises three different issues. First, is it possible that in using up one resource we create another of equal value? Secondly, how much would it cost to use up the resource more slowly? And thirdly, how much do we weight the claim of future generations to consume when compared with our own claim? In addition, in answering these questions we must distinguish between the position of any one country and the position of the world as a whole, since it is open to one country, but not at

present to the world as a whole, to exhaust its own minerals gaily, while trusting to be able to import from other places in future.

If we take first the using up of one resource in order to create another, we can see this most clearly if we consider the position of a single country. A country with a low standard of living, like Northern Rhodesia or Malaya or Trinidad, may find itself to be possessed of a mineral resource which the rest of the world values highly. If it refuses to exhaust this resource, its standard of living remains low. On the other hand, if it exploits the resource, funds become available for improving its capital equipment in other ways. More can be spent on education, on improving its agricultural land, on irrigation facilities, on public utilities, on research and on surveying to find new resources or new uses for other resources. As a result the exhaustion of its minerals may yet leave the country in a much better position to earn its living in the future than it would otherwise have been: one resource has been transformed into another. This is not, of course, always the case. Often the proceeds are used wastefully, or mainly on current consumption, so that when the resource is exhausted there is nothing left to show for it, and the economy lapses into stagnation: this is a common feature of former mining towns, and of some countries formerly rich in minerals. Often, too, the proceeds accrue to some other country; it may be foreign shareholders who reap the lion's share of the proceeds of mining, and use the funds to improve the capital equipment of their own countries, instead of the country out of whose soil the mineral is coming; or it may be that the mining is done by immigrants, who rush in, extract what they can, and emigrate when the mineral is exhausted, leaving nothing but an ugly mess behind. Mining can be made to create resources as valuable as those which are being exhausted, but this does not happen unless the country insists that a large part of the proceeds is to be invested in new resources (including education). All the same, the new resource is not always an adequate substitute for the old. Take the position of a primitive country which discovers that it possesses coal, or iron ore. These are resources on the basis of which great industries can be established. It may not be in a position to establish such industries, because its people lack the necessary education and capital. So it may decide temporarily to export its ore or its coal, as a means of earning funds with which to build up its productive capacity. Yet, if it is exporting the ore or the coal, when the time comes that it is in a position to establish industries on these bases, there may be no ore or coal left. With these two minerals in particular it is not always easy to decide whether they should be exported now, as a source of funds, or kept for some uncertain future as the basis of local industries.

Much the same difficulty arises when we move from the level of the individual country to that of the world as a whole. It is then only in a very limited sense true that in using up one resource one may create another of equal value. It is true that the rapid using up of minerals in the past two centuries has stood behind the enormous increase of our knowledge and of our productive capacity; future generations would not have been advantaged if we had bequeathed them the minerals in the ground without the knowledge of how to use them or all the other scientific knowledge which they will inherit. But what use will they find in a great deal of knowledge if they do not also get the resources to which to apply it? It is possible also that this knowledge may enable them to discover new resources, or new uses for what was previously considered useless (bauxite and uranium were until recently just types of 'stone'). It may even enable them to synthesise all they need from the air, building it up from atoms of hydrogen. In other words, it is difficult to assess just how much of a bad turn, if any, we do to our descendants by using up resources now. They may be better off, because of the knowledge and other capital which we are thus enabled to bequeath them in exchange; or alternatively they may curse us for our improvidence, just as some Middle Eastern and North African peoples may now feel inclined to curse their ancestors for destroying the forests which were once there, and whose removal has so extended the area of the desert.

The rate at which we use up resources is also determined by the cost of using resources with greater care. For example, in mining there are ores of varying degrees of purity. It is always possible to get more mineral out of a given patch of ground by taking out also the inferior ores. Similarly, forests can be cut more or less destructively, and greater or less care can be taken in replanting. Exactly the same applies to agriculture. In most countries it is becoming a moral law (sometimes even with legal sanction) that soil fertility must be conserved. This has by no means been the universal attitude. On the contrary, in many countries shifting cultivation is still exceedingly common, with its implication that it

does not matter if we exhaust the fertility of this bit of soil because we can pass on to another bit next year; the case differs slightly from that of minerals if the soil is able to regain its fertility when left fallow, but in fact many soils treated in this way have lost their mineral content, and even their capacity to rebuild their humus supplies. In all these cases the natural resource could be preserved, whether absolutely or merely to a greater extent, at some cost. The commercial user matches this cost against the gain; the cost of mining the inferior ores, against their price; the cost of replanting the forest, or of cutting more selectively: the cost of measures of soil conservation. If the community does not accept this judgement, it has ways of inducing or of enforcing more careful use of resources. Inducement takes essentially the form of subsidising methods which give more intensive use; such is the effect of the method of levying royalties on gold mines which is now used by the government of South Africa; or of subsidies for planting trees; or subsidies for bringing inferior land into cultivation, or subsidies for soil conservation measures. Alternatively, the law is called upon to enforce more careful use, by laying down standards for replanting or for soil conservation, and by imposing penalties for breaches of these standards.

Behind all this lies the question, what does our generation owe to the future? Why should we not use what we want now, and leave future generations to fend for themselves? Why should their happiness in the future be accounted of greater importance than our happiness now? For example, take the population problem. Suppose that a country has enough coal to supply 10,000 million man-years of consumption. Why is it better to have twenty million people consuming it over 500 years rather than fifty million people consuming it over 200 years? Or take the case of soil conservation; if we use up enough of our labour now for this purpose we can bequeath the soil to our descendants not only in as good but also in a much better condition than that in which we have found it. But why should we incur these costs on their behalf? Or why, for that matter, do we make any investments now whose fruit will not be fully enjoyed in our own lifetimes – investments in damming rivers for hydro-electric power, for instance? These questions are not answerable except in terms of our respect for the continuity of the human race. Most of us have a feeling – whether it is an innate instinct or something we are taught cannot be said – that the future of our community is important, and that each of us in particular, and our generation in general, ought to make some sacrifice of present comfort in the interest of generations to come. There is no means of measuring how large this sacrifice should be, and, correspondingly there is no objective answer to the question 'What is the right rate at which to use up resources?' Each community has to decide these matters for itself in each generation.

We are on firmer ground when we inquire into the relationship between current output and current population. The relationship between population and output per head is a matter of the advantages of specialisation and the economies of large-scale production on the one hand, and the diseconomies of more intensive and extensive use of natural resources on the other hand. The bigger the population the greater are the opportunities for specialisation, not only of persons, but also of firms and of industries. 'The division of labour is limited by the extent of the market.' It is true that the existence of international trade makes specialisation possible to some extent irrespective of the size of the country – indeed in another sense the smaller the country the more highly specialised it usually is. But since there are many activities which cannot be supplied by foreign trade – housing, personal services, internal transport, etc. – it remains true that there will be larger internal specialisation the greater the internal market. Besides, foreign trade has its own handicaps and instability which diminish its attraction when compared with internal trade. All this line of reasoning applies equally to the economies of large-scale production. These economies can sometimes be secured by producing for export, but in many cases (e.g. some public utilities) the product is not exportable; and in any case, because of the greater risk of foreign trade, the economies of scale are more likely to be exploited for the domestic market than if the investor has to rely on disposing of the major part of his output in foreign markets.

The industries which benefit most from a large market are the public utilities, and some factory industries which work with metals, especially in the early stages of metal production. The public utilities – transport, electricity, gas, water – show very marked economies as population grows denser, since their roadways, pipes and transmission lines are then more fully utilised. The factory industries making consumer goods and machinery on the whole reach their optimum size very soon, unless they are assem-

bly industries. It is mainly the factory industries processing the metal ores, and those making basic chemicals, which enjoy the largest economies of scale. All the same, most industries gain advantage from operating in a country which is large enough to support a range of other industries, even if the average factory is small, because of the extent to which industries depend on each other either as sources of materials, components, and services, or else as purchasers of intermediate output or of by-products. On the other side, the diseconomies of scale show up most quickly in agriculture and in mining. As population increases it becomes necessary to cultivate the less fertile lands, or to work the fertile lands more intensively, and in either case the result is diminishing returns.

It follows that the population which a country can carry without diminishing returns depends upon whether its natural resources fit it to be a manufacturer of metal goods and heavy chemicals, or whether they fit it primarily to engage in agriculture. In the former case, it may continue to secure increasing returns over a considerable increase in population, whereas in the latter case diminishing returns set in much earlier. We have also the paradox that a country may be overpopulated relatively to its agricultural resources, but underpopulated relatively to its capacities for industrial development. Some very small countries, like Jamaica or Mauritius, face the problem that their populations are much too large in relation to agriculture, and at the same time much too small to support a wide range of industrial development.

It follows also that a country is not to be described as overpopulated simply because it has more people than can be fed from its own lands. This is another sense in which the term is sometimes used. It would indeed be a useful sense if international trade in food were not possible, or were very costly, or if one were studying the problem mainly in the light of military strategy. In Western Europe it takes about one acre to feed one person, if we include children as one unit each, and if we take three acres of grassland as roughly equivalent to one acre of arable land. In the United States of America the standard of feeding is about the same, but the productivity of land is much lower, so it takes more than two acres to feed each person. The position of other countries is determined on the one hand by their standard of feeding, and on the other by productivity. A great deal depends upon the amount of livestock products consumed (meat, milk, butter, etc.), since these are very costly in terms of land. For example, India consumes less than two-thirds of the calories per head consumed in Europe, and only a fraction of the protein, but this is offset by the lower productivity of land; her population therefore feeds at the rate of about four-fifths of an acre per person.

Given the possibility of importing food, the carrying capacity of agricultural land ceases to be decisive in determining what the maximum population should be. A country may be in a position to increase its income by concentrating on developing other more valuable skills or resources, and may deliberately leave lands idle and import food, even though it could feed itself if it had to. A country which has not enough land to feed its population may yet be underpopulated if it has other resources or skills such that a further increase in population would increase output per head. This is not to say that a country cannot be overpopulated so long as it can engage in manufacturing, or in some other activity, for it is quite possible to be overpopulated in manufacturing, as well as in agriculture, or in anything else, if a smaller population would show a larger output per head. The point of this paragraph is merely that one must take all activities into account when judging whether a country is overpopulated, and not its agricultural resources only.

However, while it is true that the individual country is not necessarily overpopulated merely because it cannot feed itself, the same argument does not apply to the world as a whole. As we have seen, estimates of the current carrying capacity of the world lie between 2500 million and 10,000 million people, and at the present rate of increase ($1\frac{1}{4}$ per cent per annum) world population would reach the upper figure in just over a century. On the other hand our techniques of producing food are improving all the time. Yields per acre in the agriculturally advanced countries have in the past shown some substantial increases, and it is quite impossible to foresee what the rate of technical progress will be in the future. The fact that world population is growing faster than world yields per acre causes many people to give added weight to the argument that it is dangerous for any country to rely on being able to import food. They think, for example, that the British future would be much more secure if the population of Great Britain were not more than twenty-five million, arguing that at that level the population

would be large enough to exhaust practically all the large-scale economies of the public utilities and manufacturing industry, while at the same time small enough to reduce dependence on outside food supplies to very small proportions. All these arguments, however, necessarily rest upon speculations about a most uncertain future.

It is also necessary to make the point that even if we could establish that output per head would be bigger if population were say 20 per cent smaller, it would not follow that output per head would rise if that number emigrated, or if the birth rate fell. These population comparisons are based on the assumption of an unchanged population structure in the sense that the proportions of old and young, male and female, skilled and unskilled remain the same. Whereas, as the population changes, its structure also changes, not always for the better.

The remaining sense of the term overpopulation refers to the state of affairs where the population of a country is so large that if it were larger output would be no bigger. This is the extreme form of the first sense of the term. In the first sense output per person decreases as population increases, but total output increases; in this sense total output does not even increase. This extreme state of overpopulation is, alas, not unknown. It can usually be recognised from the excessive numbers attached to certain sectors of the economy, especially domestic service, petty trade, casual employments and agriculture. Domestic service swells because such economies adjust themselves to the need for each person to provide as much employment as he can; social prestige requires each person to have servants if he can, and the wealthier members of the community have to fill up their households with hordes of retainers who are little more than a drag on the purse. An extreme example of this is the island of Barbados, where 16 per cent of the population is recorded by the census as engaged in domestic service. Petty trading shows similar extension; market places are crowded with stalls, in which each seller makes only a very few sales – selling and passing the time away in gossip being almost indistinguishable. Then there are the hordes of porters, jobbing gardeners and others who get such casual employment as they can, averaging perhaps a day a week or less. In agriculture the phenomenon shows itself in the smallness of the farms; the plot cultivated by the average family is so small that it cannot fully occupy all the time of members of the family. Whether the excess of population shows itself mainly in agriculture, or mainly in domestic service, trading and casual jobs, depends on whether agriculture is operated with wage labour or by peasants. If it is operated with wage labour (e.g. Barbados) it will not employ more people than are needed for cultivation, and the surplus must then find a living outside agriculture. But if it is operated by peasants, the surplus lives on the family farms, and there may be very little surplus in other occupations. The normal tendency in overpopulated countries is for the great landowners to let their lands to peasants for rent, rather than to hire agricultural labourers. They get more this way, since the wage they would have to pay to labourers is more than is left to the peasants when rents have been extracted from them.

Various attempts have been made to measure overpopulation in this extreme sense. The measurement of the rural surplus depends upon estimating how many gainfully occupied persons are needed in agriculture per acre under cultivation, given the crops, the techniques, and the equipment in use. Some crops are much more labour-intensive than others; rice much more so than wheat, and sugar or tea much more so than maize or cocoa or rubber. The equipment also makes a great difference, since whereas a family with hoes is fully stretched with three or four acres, a family with draft animals and ploughs can manage ten to fifteen acres, and a family with a tractor thirty acres or more. It has also to be remembered that labour requirements vary with different operations; in some crops the maximum labour requirement is for ploughing, while with others it is for harvesting. It follows that the degree of overpopulation cannot be estimated by some generalisation, but must be calculated separately for each locality. When such calculations are made for the less developed economies, the labour requirement in cereals (excluding rice), with plough and draft animals, ranges from about fourteen to about twenty persons per 100 acres under cultivation. Detailed calculations for India, where there are currently about twenty-seven persons gainfully employed in agriculture for each 100 acres under cultivation, have led to the conclusion that at least a quarter of the agricultural population is surplus to requirements. This is equivalent to having some twenty million people permanently unemployed; hence the phenomenon is often referred to as 'disguised unemployment'. This phenomenon is rare in Africa and in Latin America, but it repeats itself in China,

in Indonesia, in Egypt and in many countries of Eastern Europe.

This sort of overpopulation, in addition to wasting labour, frequently also reduces the fertility of the soil. One reason may be that in extreme cases the people cannot afford to keep cattle, because cattle eat too much; so the land does not get manure. (The Indian peasants cannot afford to keep as many cattle as they do, but hold on to cattle for religious reasons; the land does not benefit as it might, however, since half the dung is burnt as fuel.) Another reason is the pressure to use every inch of ground; land is put under the plough which ought to be left in forest, or which should be left for purposes of soil conservation. Then there is the temptation to overcrop the land; to take off too many crops in one year, or to cut down on fallow periods. The Law of Diminishing Returns states that if there are too many people on too little land the marginal product of labour will be negative, and this, alas, is only too familiar a feature of overpopulated countries.

Given a population of this size the obvious policy is to develop as much employment as one can off the land. This is not only the obvious means of increasing non-agricultural output, but it may also be a necessary move in trying to increase soil fertility itself. If one could get some of the people off the land, return some land to forest, use some more for erosion control and increase fallow periods, agricultural output would by definition increase, though not necessarily at once. It would also be possible to increase the size of the average farmer's holding, but this would not necessarily increase agricultural output, since yields per acre tend to be highest on the smallest farms; it might increase output, however, if the farmers, now better off, were able to save more and to invest in improving their land. It is not, however, by any means easy to expand non-agricultural employments rapidly enough to cope with the increase of population and as well to reduce some of the agricultural surplus. Suppose that the farm population is 70 per cent of the whole, and that population is increasing by $1\frac{1}{2}$ per cent per annum. Then, if the rural surplus is to be reduced, non-agricultural employments must expand by more than 5 per cent per annum. Not many of the less developed economies have succeeded in industrialising rapidly enough to reduce their agricultural populations absolutely. Japan and the U.S.S.R. have managed to do this, but their rates of industrial expansion have been phenomenal, compared with say the U.S.A. or Germany.

Employing more people off the land does not provide food; on the contrary, it increases the demand for food relatively to the supply. Any such policy must therefore be accompanied by a vigorous programme for increasing yields per acre, through an extensive network of agricultural instruction, through greater use of fertilisers, through multiplication and distribution of better types of seed, through greater conservation and distribution of water and in all the other ways which Japanese experience has demonstrated to be capable of increasing agricultural output rapidly. But the policy also involves much more than this. When the standard of living is low the home demand for manufactured goods is small – less than 15 per cent of national income, in terms of income generated in manufacture (i.e. excluding the value of the materials used), and even less than this in terms of numbers employed. Hence if the people are to be employed producing manufactures there will soon come a point where more manufactures cannot be sold profitably on the home market, so that if full employment is to be attained surplus manufactures must be exported. This is the common fate of all countries which are overpopulated relatively to their agricultural resources – Britain, Japan, Egypt, Germany, India or others – they can make a living for all their people only by exporting manufactures in return for food and raw materials. The development programmes of all such countries have to provide for making a bid for capturing foreign markets in manufactures (e.g. Germany, Japan), or else (like India's 'First Five-Year Plan') they have to play down industrialisation, and to leave the rural surplus where it is.

Capturing a larger share of the world market for manufactures is no easy task. The market for metal goods and engineering products generally increases steadily, but the market for other goods grows very slowly, or in some cases (e.g. textiles) contracts absolutely. Hence the countries which are well endowed with coal and ores can increase their shares of world trade in manufactures relatively easily if they try to. On the other hand, those overpopulated countries which have very little coal or metallic ores can compete effectively only in the types of goods (textiles, leather goods, finished small wares) which are increasingly difficult to sell. To say this, however, is only to say that a country which allows its population to grow excessively in relation to its natural resources will have very great difficulty in providing full employment for

its people, and in assuring them a reasonable standard of living.

Is it possible, in the light of this discussion, to arrive at conclusions on the present degrees of under- or overpopulation of the various countries of the world? This is an extremely hazardous exercise, since the resources of the various countries are not known, and since the potentialities of what is known change with new techniques and new demands. However, for what it is worth we may hazard the following guesses about the continental masses. Africa is underpopulated, in the sense that that continent still has empty cultivable lands, and in the sense that its present sparse population makes the provision of public utilities very expensive; if Africa had a larger population the cost per head of roads, electric power, water supplies, railways, hospitals and other services would fall, while the quality could also be improved. There are some congested areas in Africa, in Eastern Nigeria, in parts of Kenya, and in parts of the Union of South Africa, but the continent as a whole, south of the Sahara, is underpopulated. Latin America and Australia are probably both underpopulated in the same sense, though the extent of the empty cultivable lands is in both cases much more uncertain. At the other extreme, Asia is heavily overpopulated, though there are parts of South-East Asia of which this is not true. Time may alter the value of the natural resources of Asia; new techniques may make her soils much more productive than we at present imagine possible, or great new mineral resources at present unsuspected may become available. On current knowledge, however, there is no doubt that the number of mouths there are to feed keeps down the standard of living of Asia. In between come the economies of Europe and of North America. These can be complementary, and should be considered together. Europe is not overpopulated if she can get food and raw materials cheaply from North America, as was the case until 1939. Canada may be underpopulated, having regard to her newly discovered resources. The United States is a very doubtful case, probably neither over- nor underpopulated. The tentativeness of judgements relating to Europe and North America emphasises how difficult it is to weigh in the balance the diseconomies of a large population with regard to food and raw materials on the one hand, and its economies in regard to manufacturing, public utilities and other services on the other hand.

Previous extracts by Ryder, Geddes and Learmonth have already indicated some of the social implications of population change. These are here explored further on the basis of two particular issues. First, the article by Ronald Freedman points out that population change is itself, in part, the result of a changing social situation. In underdeveloped as well as developed countries there are social rules which affect individual couples and hence the overall fertility pattern. These social rules are, in turn, influenced by a number of other institutions in society – illustrating once again the complexity of social reality and indicating that change in the present high fertility patterns of many developing nations may be impossible without simultaneous change in a number of other social institutions.

In countries with high population growth there has often been migration from the countryside to the towns. Such migration results from a variety of factors of which population increase is one, and the article by Myron Weiner considers the political effects. The two extracts are only illustrative of the wide-ranging significance of changes in population; these changes are part of a complex process of social and political change in which population growth can be analysed as both cause and consequence. They show at the same time how a social scientist can deepen our understanding of a particular social phenomenon and in the process indicate, as Weiner does, the falsity of an accepted, commonsense explanation.

NORMS FOR FAMILY SIZE IN UNDER-DEVELOPED AREAS

Ronald Freedman

It is commonplace observation that fertility is high in the so-called underdeveloped societies. Sometimes this leads to the erroneous view that their fertility is limited only by what is biologically possible so that birth rates reach a mythologically high natural level. In this idealised view free and unrestrained sexual unions beginning at an early age produce very high fertility as an incidental result without deliberate intent and without much individual or social concern about family size.

A more plausible general proposition is that reproduction, whether at high or low levels, is so important to the family and to society everywhere that its level is more or less controlled by cultural norms about family size and such related matters as marriage, timing of intercourse and abortion. In each society the cultural norms about these vital matters are consistent with social institutions in which they are deeply embedded. Changes in fertility are unlikely without prior or, at least, simultaneous changes in these institutions.

Before developing these ideas, it may be useful to consider briefly the variability in fertility among underdeveloped societies. A recent study of the Cocos Island population[1] provides a unique example of exceptionally high fertility expected in the model of virtually unrestrained fertility. The Cocos Islanders, living between 1870 and 1947 under an exceptionally favourable paternalistic regime, married early, and about two-thirds of their brides were pregnant at marriage. The birth rates between 1880 and 1947 averaged about 55 per thousand and reached 60 in several years. This may be contrasted with birth rates of about 35 per thousand in pre-industrial England. The gross reproduction rate for the Cocos Islands was about 4·2, perhaps the highest ever recorded with reliable data.

This is an exceptionally high fertility level. It is not characteristic of all underdeveloped areas, although it is not unlikely that there are many populations with a biological potential for a similar performance. Some African populations may reach similar levels.[2] In Appendix I, estimated gross reproduction rates are listed for a number of high-fertility populations, underdeveloped by most criteria. These reproduction rates are very high in comparison with the rates for the developed low-fertility countries shown for contrast. But the rates generally are well below the Cocos rate or that for the American Hutterites. The rates do not conform to any uniform level.

The range of fertility rates currently found among underdeveloped areas is wide enough to include rates which characterised Western Europe before its modern fertility decline but long

after social and economic development were under way. For example the gross reproduction rate of 2·4 for England in 1861, shortly before fertility began to decline, is lower than the rates for most but by no means all of the high-fertility societies today. Married women born in England in 1841–5 had just under six children on the average,[3] if they survived to the end of the child-bearing period. This is about the same as the figure reported for India now,[4] although it is significantly less than the average of more than eight children born to the Cocos Island mothers and the average of more than ten to the American Hutterite mothers.[5]

These selected data are intended only to illustrate the idea that fertility, however measured, varies from moderately to very high levels in underdeveloped societies, even after allowing for inaccuracies of base data.

Genetic differentials in fecundity eventually may be shown to account for some of the existing variation between societies, but I think they are unlikely to explain the major variations in time or between societies. They are unlikely to explain why fertility is relatively high in underdeveloped societies and declines with development.

As a more plausible explanation, I begin with the thesis that societal levels of fertility are related to variations in cultural norms about reproduction, and these, in turn, are related to the nature of the society.

One of the fundamental principles of sociology is that when many members of a society face a recurrent common problem with important social consequences they tend to develop a normative solution for it. This solution, a set of rules for behaviour in a particular situation, becomes part of the culture, and the society indoctrinates its members to conform more or less closely to the norms by explicit or implicit rewards and punishments.

The problem of how many children a couple should have is so widely shared and has so many personal and social consequences that it would be a sociological anomaly if social norms regarding it did not develop. This is also true for such related problems as when to marry and when sexual intercourse is permitted. In view of the special importance attached to kinship ties in the underdeveloped societies, it would be particularly strange if the reproductive level of the familial unit were not a matter of normative concern.

Norms about family size are likely to be in terms of a range in numbers of children that are permissible or desirable. While specifying clearly that childlessness is an unspeakable tragedy and an only child very undesirable, the norm for a particular culture or group may be as vague as 'at least three or four children' or 'as many as possible'. But I know of no organised society, primitive or modern, in which the question of how many children are born is a matter of indifference either to the reproducing unit or to the community.

In various underdeveloped societies, a large number of sample interview surveys[6] have been made since World War II in which peasant populations have been asked such questions as how many children they want for themselves, how many are right for others, etc. In almost all of these studies only a small minority of the respondents found the questions ridiculous or meaningless or answered that such matters were up to fate or God, etc. Even where the answers invoke the Deity or fate, further inquiry frequently indicates that the respondent has rather firm ideas of what he desires that fate or divine providence should bring. In most such surveys the answers tend to be clustered about a moderate modal value rather than to be randomly distributed. Appendix II contains some illustrative frequency distributions of attitudes about family size and family growth from surveys in several different countries. Of special significance are data in a number of these studies indicating that a majority of those with three or four children want no more and many prefer a smaller number than they have. (Illustrations are found in Appendix II.)

There are a few studies in which it is reported that questions about the desired number of children are regarded by respondents as ludicrous. This is reported, for example, by Richards and Reining for the Bahaya of Africa.[7] But the attitude expressed apparently does not result from the absence of a norm. The report is that, under existing conditions, most women do not have the four to six children in the household which they consider ideal and would approximate if they could. And even in this instance, the authors report that there are voluntary practices limiting reproduction during certain periods.

The exact meaning and accuracy of the answers obtained in such surveys is open to question, but there is little doubt that there

is a considerable consensus in such populations about desirable family size.

It must be admitted that much of the survey evidence is from societies that have been in touch with the West for many years, but some of the samples studied have been in village areas of India, the Near East and other places where illiteracy and immobility have isolated much of the population from modern ideas.

It is possible, but unlikely, that all of the populations studied through sample surveys have already been considerably influenced by modern Western values about reproduction. The same cannot be said for a large number of primitive societies whose ethnology with respect to sexual and reproductive behaviour was studied by Ford.[8] He concluded that childbirth is by no means universally accepted as either natural or inevitable. Many women are ambivalent about childbirth, so that considerable social pressure may be necessary to ensure adequate reproduction.

Obviously, social norms encouraging reproduction at less than the physiological maximum can be implemented only if one or more practices for fertility reduction is fairly widespread in the population, so it is very relevant to ask what control practices are available. A variety of controls are possible and are used. In addition to such deliberate and obvious practices as contraception and abortion, there are a large number of others; for example, the age at first sexual union, timing and frequency of intercourse, voluntary or involuntary fecundity impairments, foetal mortality, etc.

Davis and Blake[9] have provided a useful classification of such control factors immediately determining the fertility level of a society. They call them 'intermediate variables' because they are the means of fertility control standing intermediate between the social institutions and norms on the one hand and actual fertility on the other. Any social influences on fertility can only operate by affecting one or more of these 'intermediate variables'. The Davis and Blake classification is shown in Appendix III. I will not duplicate the detailed discussion of these variables given in their excellent paper.

Different combinations of values for these intermediate variables may produce identical fertility levels. On the other hand, societies with very different fertility levels may have similar values on some though, of course, not all of the intermediate variables. It is unnecessary and often incorrect to assume that levels of these control variables always are manipulated deliberately to limit fertility. The limiting effect on fertility is often, probably usually, an unintended consequence of one or more cultural patterns that have no explicit connection with fertility.

Contemporary data for high-fertility societies make it evident that a combination of very different control variables may produce high fertility. Coale and Tye[10] have recently shown that high fertility in some Chinese populations results from late marriage followed by very high reproduction rates into the later childbearing years. On the other hand, high fertility in India results from very early marriage and high reproduction rates in the early childbearing years, with a sharp decline in the later years. While both types of controls may lead to similar average family size, the 'Chinese' pattern will result in a slower rate of population growth, other things being equal, because the length of a generation is greater.

Cultural norms controlling premarital intercourse and the age at marriage have been particularly important in reducing fertility levels in a number of pre-modern societies. Apparently, this explains the fact that fertility was only moderately high in pre-industrial Europe. Economic historians[11] have shown that property and labour arrangements in a number of West European areas encouraged celibacy or the postponement of marriage in order to maintain certain standards of what the economic unit should be, and of what a couple should have in order to marry and raise children. Together with controls on premarital intercourse such arrangements apparently kept fertility below the levels of other pre-industrial societies. The breakdown of these controls in the first stages of modern urban-industrial development apparently resulted in an increase in fertility in some places before the long-run decline in fertility set in. The fact that fertility was only moderately high in many areas of pre-industrial Europe is evidence that very high fertility is not a universal characteristic of such societies.

Davis and Blake point out that since having some children is very important in pre-industrial societies, a society with high and variable mortality often is likely to have built into its structure strong pressures for having children early in marriage before one

or both of the parents die and also for having some 'extra' children as a safeguard against the catastrophic loss of the essential minimum number. If unfavourable economic conditions develop, this may result in 'too many' children. Therefore, there is likely to be a delicate balance of pressures towards higher fertility to ensure at least a certain minimum number of children, and counter-pressures to minimise an intolerable surplus of children under difficult subsistence conditions. Davis and Blake present the hypothesis that this necessary pressure in opposite directions accounts for the adoption in many pre-industrial societies of those control practices which ensure a minimum fertility level but permit a reduction in the numbers of children late in the childbearing process, principally abortion and infanticide. These counter-balancing pressures also may account for the failure to adopt contraception, which operates very early in the reproductive process.

It would be desirable to have a systematic classification of societies by the values of the intermediate variables. Then we could study how particular combinations of the control variables produce a particular fertility level, and how each pattern is related to social norms about family size and ultimately to the social structure. Unfortunately, the data for such systematic classifications are not available.

However, Carr-Saunders,[12] Himes,[13] Ford,[14] Deveraux[15] and others have assembled scattered evidence on the existence in many pre-industrial societies of a wide variety of fertility control measures including contraception, abstinence, abortion, infanticide and delayed marriage. In many cases the limiting practice is not perceived as having birth limitation as an objective. Most of the evidence demonstrates the existence of certain practices in a culture without specifying either the extent of use or the effect on fertility. Nevertheless, there is basis for the supposition that more or less effective methods of population control potentially were available in many pre-industrial societies and that this probably affected fertility levels in most of them. I venture the view that such practices might have had much wider use and greater effect were it not for the rewards derived from having children in such societies and the risk that these rewards would be lost because of unpredictably high mortality.

Contraception is not one of the potential control practices widely prevalent in any large underdeveloped society. An important unresolved question is whether this has been because adequate contraceptive technology was lacking or because normative pressures for high fertility precluded either the development of new methods or the adoption and dissemination of known methods beyond a small minority of the population.

Despite evidence that various types of contraception have been known and used by small numbers in many pre-industrial societies, there are very few examples of their adoption by sufficient numbers to affect the general fertility level. Coitus interruptus, requiring no mechanical or chemical materials, has been known in many societies for a long time. Potentially this practice might have diffused to a wider group from the small minority using it in pre-industrial societies with resultant impact on fertility levels. It is difficult to dismiss coitus interruptus as unsuitable for mass adoption, since this was the principal means by which the British and French drastically reduced their birth rates in the modern period.[16] Coitus interruptus is still one of the most important methods producing the relatively low birth rate in Britain.[17]

Many hypotheses have been advanced to explain why coitus interruptus and other possible methods were not adopted or why alternate methods were not developed before the modern period. Among various explanations, two emphasise the internal structure of the family. The combination of male dominance and the lack of communication between husband and wife, especially on problems of family size and family planning, is said to retard the adoption of family planning. This might be especially true if, as has been suggested by some, the wife has the stronger interest in family limitation, while the husband is indifferent. Dominance of the husband combined with his indifference on this issue might be particularly important for such a male-dominated method as coitus interruptus. Hill, Stycos and Back[18] have provided some evidence from Puerto Rico that customary barriers to husband–wife communication may be related to failure to adopt family planning in Puerto Rico now.

Whether this type of explanation is valid is an open question. There is a *prima facie* case for the interest of the father, too, in limiting fertility in a subsistence economy, since children are consumers as well as producers. Some recent studies in underdeveloped areas indicate that the father's disinterest may have

been exaggerated. Studies in India, Ceylon, Taiwan, Puerto Rico and elsewhere[19] indicate roughly similar attitudes on the part of husbands and wives on many issues about reproduction. If it is true that pre-industrial couples do not discuss fertility control or family size, this may be because a long-standing consensus on high fertility leaves little to discuss.

Other explanations for lack of use of contraception in some cultures centre on the idea that modern contraceptives were not available and that coitus interruptus was not acceptable in many cultures. Whether it was simply the availability of contraceptive means which determined fertility levels historically is currently of great practical importance. If fertility levels were high in pre-industrial societies simply because of inadequate contraceptive technology, then current programmes for reducing fertility in such areas can concentrate on making modern contraceptives available. This very simple solution is no longer widely espoused, because programmes for making contraceptives available in underdeveloped areas have not been very successful to date. While it is true that no country as yet has mounted an efficient all-out campaign to disseminate contraceptive supplies and information, the efforts made have been intensive enough in some places to indicate that the motivations to limit family size either are not yet very powerful or are restrained by powerful cultural counter-pressures.

While the issue is certainly very controversial, I take the position that a variety of control measures, including some forms of contraception, have been available potentially in underdeveloped areas and that past failure to use them more extensively has been a result of normative pressures for high fertility. On the individual psychological level this reduces to a simple statement that couples had many children because they wanted them, not because they were ignorant of how to avoid having them. This does not preclude the probability that ignorance about control methods was prevalent. Nor is it necessary to assume that the size of family achieved was deliberately planned. It does require an assumption that the values about family size and limiting practices, deliberate or not, were in rough correspondence.

Obviously, at the proper time and place provision of suitable contraceptive supplies and information is indispensable for fertility decline. I am questioning only whether this necessary cause is also a sufficient cause for a fertility decline.

Up to this point, I have been concerned mainly with indicating that control measures were available potentially in the underdeveloped areas so that their absence is not the explanation for high fertility. But to support the position that high fertility was normative, it is also necessary to consider why there have been normative pressures for large numbers of births.

I believe that most demographers and sociologists would agree on two very general explanations. From either the individual or the social point of view, high fertility has been an adjustment both to high and variable mortality and to the central importance in community life of familial and kinship ties.

In most pre-industrial societies a wide range of activities involve interdependence with kinsmen and especially with children. These include production, consumption, leisure activity, assistance in illness and old age and many other activities covered by non-familial institutions in modern societies. To simplify greatly: large numbers of children are desired if the values considered worth while are obtained through familial ties rather than through other social institutions. If kinship ties are very important in a society where mortality is high and variable, the number of births desired and produced will be especially high in order to ensure the survival to adulthood of the essential minimum number of children. Selective experience favours the development of beliefs and practices encouraging fertility high enough to minimise the grave risks of few or no surviving children.

Because mortality is such an important consideration in underdeveloped areas, it is necessary to distinguish between norms for large numbers of births and norms for large numbers of children. A considerable number of studies in contemporary underdeveloped areas have found that three or four children are reported as desired by the populations studied. This is contrasted with the six or more births women actually have had by the end of the childbearing period. A frequent interpretation of these results is that the number of children desired has sharply declined from higher earlier levels, evidence of a trend to modern small-family values. But it is quite possible that the number of living children desired really is not much lower now than it was a hundred or more years ago. With high mortality, six or more births are required if three or four children are to survive.

For example, a United Nations study in Mysore, India, in 1952[20]

found that for a sample of living women about forty-five years of age, only about 66 per cent of the children they had ever borne were still alive. This figure exaggerates somewhat the percentage of children surviving, since there is evidence of under-reporting of children born alive but dead at the time of the survey. In the Mysore study, the average number of children considered ideal by respondents is fairly close to the average number surviving to women living to the end of the childbearing period, although it is much lower than the number actually born.

In India and elsewhere the finding that three or four children are considered desirable frequently is considered paradoxical, since the populations surveyed make little use of family planning information and supplies when these are available. However, wanting three or four living children in a high-mortality country is quite consistent with the much higher average number of births. Even if mortality is declining, the peasant who has learned from his culture to depend on his children for labour on the farm, for old-age security and for other essentials cannot be expected to extrapolate declining mortality with the demographer and to calculate a long-range need for fewer children. Significantly, several studies indicate that favourable attitudes to the practice of family limitation are much more common among those who already have three or four living children. This is consistent with the view that the availability of means for birth control is largely irrelevant until what is regarded as the essential minimum number of children is secure.

This suggests that known low mortality is one of the necessary conditions for an effective social policy for reducing fertility. Historically developed normative pressures for large numbers of births may persist for a long time after mortality has fallen, since they are closely tied to many aspects of the social structure. Housing allocation and other social arrangements in high-fertility countries are developed on the implicit assumption that many of the family members, and especially many infants, will die. When mortality declines, increasing pressure is exerted on many traditional social arrangements, thus creating a ferment for change. But this takes time, and historically there has been almost always a considerable lag between the fall of mortality and the fall of fertility. It is likely to be particularly difficult to counter the traditional pressures for high fertility when the mortality decline is only prospective. This is what is being attempted in many underdeveloped areas today.

Compensation for high mortality does not account alone for the differences in fertility between developed and underdeveloped areas. After taking into account the effects of mortality, the number of children desired probably is still somewhat larger, on the average, than in the more developed societies.

Why does the social norm prescribe a relatively large number of *living* children in underdeveloped countries? In seeking an answer, I begin with the premise that the norm depends on how having a particular number of children affects the ability of the familial unit to attain socially valued goals. This, in turn, depends on the division of labour between the family and other social institutions and on how much the performance of important functions by the family depends on the number of children produced in it. The assumption is that family-size norms will tend to correspond to a number which maximises the net utility to be derived from having children in that society. Obviously, different aspects of the society may exert opposing pressures on the norms, so a balance must be struck. Therefore, we must look for important aspects of the social organisation which support the norms for family size by providing explicit or implicit social rewards or penalties depending on the number of children.

As already noted, the familial and kinship units are so important in underdeveloped societies that we may expect relatively high fertility in order to maintain them. This kind of global explanatory hypothesis lacks the precision to explain the existing and historical diversity ranging from moderately to very high fertility. But if the general hypothesis is valid, then specific variations in kinship structure should help to explain the observed variations in fertility. Anthropological research reveals a rich diversity of kinship organisation and functions in different societies. Unfortunately, systematic study of how these variations are related to variations in fertility, normative or actual, is not very advanced. Nevertheless, some broad speculative generalisations can be considered.

Corporate kinship systems involving an immortal clan with a patrilocal or matrilocal basis and affecting every aspect of life, even political organisation and war, are seen by Lorimer[21] as greatly enhancing the valuation of children. Joint family systems com-

bining several nuclear units but with periodic fission and reconstitution of the joint family make for somewhat lower fertility, since they encompass social relations less fully than the corporate system. Systems based on nuclear, neo-local familial units make for still lower fertility, since fewer routine functions are carried out within the family unit.

These broad generalisations greatly oversimplify a complex subject. Both the theory and the evidence are in a rudimentary state. Lorimer has placed emphasis on kinship organisation and especially on the role of corporate kinship systems in producing high fertility in African societies. But his work gives little attention to how the corporate kinship system affects the intermediate variables so as to produce the high fertility. Davis and Blake[22] illustrate how variations in the kinship structure operate through particular intermediate variables to affect fertility. For example, they analyse the probably functional significance of the joint family system in India in inhibiting widow remarriage and, thus, reducing fertility. The idea that neo-local nuclear family systems may lead to relatively lower fertility levels even in pre-industrial societies has been discussed mainly with reference to pre-industrial Europe. The discussion particularly has centred on various economic arrangements leading to late marriage or non-marriage, as already mentioned.

Studies in India about the comparative fertility of couples living in nuclear and joint family units yield contradictory results, with some studies indicating that couples living in nuclear units have fertility at least as high as the fertility of couples in joint households.[23] One possible explanation is that high fertility increases the likelihood that joint households will split up. But such comparisons within a society are unlikely to provide relevant evidence of how a dominant family system affects the society's reproduction level. For this purpose, the units of comparison must be whole societies or societies at different time periods.

Recent Japanese data provide a striking illustration of the close relation between norms about dependence on adult children and the course of fertility. Between 1950 and 1961 the Japanese birth rate fell spectacularly from 28 to 17 per thousand. In the same period the biennial sample surveys by the Mainichi press[24] posed to a representative cross-section of the population the question: 'Do you expect to depend on your children in your old age?' In 1950 a majority, more than 55 per cent, answered 'definitely yes'. The proportion giving this answer declined steadily in five succeeding surveys, reaching 27 per cent by 1961. It is rare that public opinion on a matter this vital changes so steadily and rapidly and just as rare that we have statistical data with which to document the trend.

These are only illustrative of theories and data about how kinship is related to fertility. No one has yet been able to assemble more than illustrative evidence relating levels and determinants of fertility under varying kinship systems, but it is unlikely that significant general explanations will emerge which do not take into account the role of kinship in the society.

This is not to deny that social norms about family size may be affected by non-familial aspects of the social organisation. For example, emphasis in the religious system on the importance of a male heir for ritual purposes probably produces pressures over time in a high-mortality society for at least two sons to guarantee the survival of at least one. This means that on the average each family will want at least four children. Such values, based on long tradition and embedded in many aspects of the society, persist long after mortality has fallen.

Taiwan is an interesting case in point of the persistence of such traditional values after mortality has fallen and even after there is an apparent willingness to adopt family planning. Taiwan has a history of slowly but steadily declining mortality under Japanese rule, followed by very rapid mortality decline after the war. Average life expectancy is now more than sixty years. The probability has been high for some time that almost all children will survive to adulthood. Fertility has begun to fall according to statistics for the whole island. In several recent small-scale pilot surveys, both husbands and wives expressed strong approval of the idea of family planning; a significant minority had begun to practise it in some form and a large majority of those interviewed indicated a desire to do so in the future. But none of this seems to portend a disavowal of traditional Chinese values. True, there is a strong consensus of the desirability of three or four children and the undesirability of larger numbers. But the preference for a moderate number of children, the number formerly achieved with larger numbers of deaths, is qualified by a strong traditional preference that at least one and preferably two or more of the

children should be sons. Couples seeking help under the Pre-pregnancy Health Programme of the Taiwanese Provincial Health Department almost always are those with at least three or four children, but they also are almost always couples with at least two sons. Strong approval is expressed in the surveys for such traditional Chinese values as the joint family, the importance of a male heir, and responsibility of children for their older parents.

Under present mortality conditions in Taiwan a single son is very likely to survive to adulthood, so having additional sons as 'insurance' is probably an anachronism, but the traditional preference for several sons persists. The family size and sex composition being sought are approximately what a traditional Chinese family might have achieved with luck under high mortality and high fertility, but fewer births are needed now to achieve this goal. The present goal of many Taiwanese appears to be to maintain traditional values in the face of the pressures of higher survival rates by using family planning.

While a preference for sons is frequently embedded in religious institutions, it may have its origin in the value of sons for agricultural labour or for other purposes. The religious sanction may sanctify and ensure what has been needed traditionally for other essential purposes.

Religious or other emphases on the importance of sons is only one of a large number of values that may affect the norms for family size, either indirectly or directly in relation to the role of the familial unit in the society. All of these must be taken into account in applying the general principle that the additional children are desired if they provide rewards greater than their costs under the particular familial arrangements of that society.

In seeking explanations for high fertility in supportive institutions and values, it would be grossly incorrect to suggest that the norms about the intermediate control variables are finely contrived in every society to produce results consistent with the norms about family size. On balance and in the long run a rough consistency is probable. But discrepancies between ends and means may develop and persist when aspects of the social structure not even recognised as related to fertility affect the reproductive level by their effect on the intermediate variables. For example, many religious systems prescribe periods of abstinence which usually (though not always) will have the effect of reducing fertility.

Illustrative of another 'unintended restraint' on fertility is the possible reduction in fecundity resulting from poor nutrition and poor health in the whole society or a stratum of it. For example, several recent Indian studies find relatively low fertility for the poorest social stratum. It is plausible that this may be a result of poor health and nutrition conditions in the lower status groups, although other explanations are possible.

A variety of such unintended effects of cultural factors in pre-industrial societies contribute to keeping fertility below its maximum biological potential. We really do not know, from systematic evidence, to what extent these operate selectively to keep fertility near the normative values. It is only a reasonable assumption that, if norms about family size are very important, the intermediate control variables will tend to reach a balance consistent with these norms. But where social change is very rapid, such a long-run balance may not be struck exactly and for short periods a large discrepancy between means and ends is probable.

Another approach to finding the variables that supported early high fertility is to study the correlates of the declines in fertility that have already occurred in developed societies. First of all, it is clear that in almost every case a substantial decline in mortality preceded fertility decline.[25] In Japan the period of lag was much shorter than in Western Europe. The lag may grow still shorter if social policy succeeds in accelerating fertility decline in such countries as Pakistan, India, Egypt and Korea. But in all the countries that have such policies at present, mortality already has declined for several decades, with little evidence of fertility decline.

Apart from lower mortality, what social and economic variables are associated with fertility decline? Given the preceding analysis there is no surprise in the plausible hypothesis that a shift of functions from family to other specialised institutions is important. There is illustrative evidence that this shift has decreased the value and increased the costs of more than a small number of children for attaining the goals men seek in the more developed society. But the demonstrable changes in family functions are difficult to disentangle from the larger complex of social and economic changes in the course of social development. Broadly speaking, development is associated with many variables, e.g. urbanisation,

industrialisation, more complex technology, greater capital investment per worker, higher living standards, and greater literacy. There is an increase in the mass media of communication and in transportation facilities which go with larger markets and the linkage of local populations in larger political, social and economic units. In these larger units, a more complex division of labour exists not only with respect to economic life but also in almost every other aspect. Thus, the tasks of an urban society are divided among a much larger number of institutions. As the economic unit ceases to be the family, recruitment of workers is less closely linked to reproduction.

Because the development process is so complex, I have postponed any definition of development until this late point. At least with reference to fertility decline, I suggest that what is essential in the development process is the shift from major dependence on relatively self-contained local institutions, to dependence upon larger social, economic and political units. Such a shift implies a change in the division of labour from one in which the kinship unit is necessarily central to a larger complex in which such local units as family and village give up many functions to larger, non-familial specialised units. In such a shift, greater literacy and the development of effective communication networks are essential.

I have purposely avoided proposing industrialisation and urbanisation as the essential developmental changes for lower fertility, although they are usually part of the change process, and I do not minimise their importance. But I think that the essential change is the expansion of the unit within which most social and economic interchange occurs.

Some evidence that urbanisation and industrialisation alone may not provide the explanation for fertility decline appears in some unpublished work by Knodel and Tangoantiang at the Princeton Office of Population Research.

Tracing fertility declines from earlier times to the present, the Princeton group found that in all of the countries in Europe, with the exception of Albania, there has been a decline in fertility of 50 per cent or more and that, with the further exception of France, the decline began some time between 1860 and 1920. The timing and steepness of the decline is not conspicuously associated with proportions of the labour force engaged in agriculture or changes in these proportions, nor is it associated consistently with urbanisation. The only index examined whose movement seems almost universally to parallel that of fertility measures is illiteracy. The level of illiteracy varied widely between countries at the beginning of the decline, ranging from 24 per cent reported in England to 79 per cent for Russia. Yet in each case, fertility and illiteracy declined together. A striking example of the unexpected lack of association between industrialisation as normally visualised and declining fertility is that the decline began in the 1880s in Italy, Hungary and Finland, as well as in England and Wales. Literacy and mass education appear to deserve special attention in relation to changing fertility, along with the more frequently stressed urbanisation and industrialisation.

How can literacy and broad educational gains affect fertility? One indirect way is by helping to reduce mortality which will later reduce fertility in ways already indicated. Obviously, literacy also will facilitate the dissemination of information about the idea and means of family limitation. But I think the role of education and literacy is more basic. I suggest that with increased education and literacy the population becomes involved with the ideas and institutions of a larger modern culture. If the individual is, or believes he is, part of a larger non-familial system, he begins to find rewards in social relationships for which large numbers of children may be irrelevant. If this thesis is correct, major expenditures for education in a development programme are justified not only for developing worker skills but also for their potential effect on the fertility level, if lower fertility is a social objective.

It is pertinent that in every country for which we have empirical evidence the spread of family planning practices to attain a smaller family has been associated with a literate audience influenced by the mass media and by a person-to-person communication of their message that transcends local boundaries. In no case up to now has a government or private programme designed specifically to disseminate family planning or small family values been the principal agency for information and change.

In the pilot study in Taiwan, referred to earlier, education and contacts with the mass media were more closely related to the actual practice of family planning than were occupation, income or rural background (see Appendix IV). It is relevant that Tai-

wan, a high-fertility country in which a significant fertility decline has begun and where just now the probability of fertility decline seems great, has a relatively high literacy rate and also has had a rapid expansion in the circulation of the mass media and in the use of the mails for personal and business correspondence. Indicative of the possible importance of wider communication is the symbolic, if humourous, fact that long-distance calls per capita are one of the best predictors of the 1961 fertility level for the twenty-two local administrative units in Taiwan.

I do not believe that education about fertility and family planning alone, completely in advance of other changes in the society, can be very effective for changing fertility norms and behaviour. Such a single specialised educational programme alone does not affect the linkage of the local population to the larger units in the broad and continuing way which can lead to the essential growing dependence on non-local and non-familial institutions. General education and literacy probably are required to do this.

High fertility norms and behaviour are too deeply embedded in traditional, emotionally-supported institutions – especially the familial – to be affected much by education or informational programmes centred on fertility alone. This is not to deny the potential importance of such programmes, once development is under way. Probably, reduced mortality sets the stage, providing the minimal threshold level required before fertility will drop. But, in addition, there must be at least minimal changes in the institutions which motivate high fertility by rewarding parents of relatively large numbers of children. There must be reason to believe that those who have smaller families can meet their needs in ways differing from the traditional patterns.

As Professor Glass[26] has aptly put this matter: 'The establishment of new demographic patterns involves the development of new incentives which press upon reproductive behaviour. Such incentives need to be embedded in a social framework. They can be encouraged by persuasion but can hardly be created by an apparatus of symbols. Symbols are important. . . . But in countries with low levels of living, perhaps even more than elsewhere, it is necessary to have something material to be symbolic about.'

This view does not preclude the possibility that broad gains in literacy and education can facilitate lower fertility levels in advance of some other changes in the society. The extent to which this is possible and what associated changes are required is still an open question. Significant studies with careful controls are possible in some of the countries introducing family planning to test whether particular levels of literacy or of other development indices are a necessary threshold to a fertility decline.

The existence of the mass media and of postal networks already involves, to some extent, the larger interaction system of which I am speaking. If the literate regard themselves as members of a larger and less parochial society, the adoption of behaviours appropriate to that society may be accelerated, if at least minimal threshold changes in society make their new self-image plausible.

Obviously, changes in literacy, education and communication networks are related to many other aspects of development. Changes in all of them probably are related in interaction to fertility decline. But it is unnecessary to assume that all aspects of development must move together. We know that there have been uneven rates of movement in the past. From a policy point of view it is desirable to look for some elements subject to social control which might lead the development process and decrease the time-gap between mortality and fertility decline. I have emphasised education because it might be such a leading variable and it is given less attention usually than economic variables or urbanisation.

To recapitulate, I have reviewed briefly the evidence that social norms support the moderately to very high fertility found in the so-called underdeveloped societies. I have suggested that high fertility and high fertility norms are not a result of unrestrained maximum fecundity but rather are an adjustment mainly to the high mortality and to dependence on kinship-based local institutions. Variations in the kinship structure were seen as a major possible explanation of variations from moderate to very high fertility. The net effect of a large number of possible control variables determines the extent to which the social norms about fertility are approximated in behaviour. Finally, I have suggested that the complex continuum from underdeveloped to developed societies is best represented for our purpose by a continuum from major dependence on relatively small local units to increasing interdependence in larger social units in which kinship plays a decreasing part. In this shift to larger units of interdependence,

education and literacy not only have an important part to play but may lead other elements in their effect on fertility.

APPENDIX I. ESTIMATED GROSS REPRODUCTION RATES FOR SELECTED POPULATIONS AND PERIODS

1. Gross reproduction rates in selected high-fertility populations

(Statistical resources are usually poor in the high-fertility areas so data are often unavailable or of questionable reliability. Those presented below are for areas where special fertility investigations, adjustment by competent investigators or unusually good official statistics make the data appear reasonably reliable. Nevertheless, these rates are less reliable than those for the low-fertility countries. They are cited to indicate the probable diversity in reproduction levels rather than as accurate measures in each individual case.)

Africa:[27]

Centre-Oubangui (1959)	2·0	Mali (1957–61)	3·3
Dahomey (1961)	3·2	Upper Volta (1960)	3·2
Guinea (1954–5)	2·9	North Cameroons	2·3
Ivory Coast (1958)	2·4		

Asia:

Ceylon[28] (1952)	2·2	Malaysian population of Malaya[31] (1946–8)	2·7
Cocos Islands[29]	4·2	Taiwan[32]	2·7
India[30] (1956)	2·6		

West Indies:[33]

Jamaica (1943)	1·8	British Guiana (1946)	2·4
Barbados (1946)	1·9	Trinidad (1946)	2·4
American Hutterites[34] (1952)	4·0		

2. Gross reproduction rates in selected low-fertility areas[35]

England and Wales:	1861	2·4	Belgium: 1928–32		1·0
	1901	1·7	1959		1·3
	1940	0·8	Norway: 1932–5		0·9
	1958	1·2	1958		1·4
Austria: 1931–2		0·9	United States: 1930–5		1·1
1959		1·3	1959		1·8

APPENDIX II. SOME ILLUSTRATIVE RESULTS FROM SAMPLE SURVEYS DEALING WITH NUMBER OF CHILDREN WANTED OR CONSIDERED IDEAL IN UNDERDEVELOPED SOCIETIES

1. Puerto Rico[36]

A. Answers to question 'Suppose you were to get married again for the first time, how many children would you want to have?':

Number of children wanted	*Sample of clinic out-patients* (%)	*General field sample* (%)
0 or 1	16	8
2	40	43
3	18	26
4 or more	10	18
All God sends	16	5
Total	100	100

2. India

A. Number of children considered ideal in Mysore area:[37]

	Mean no. considered ideal by	
	Wife	*Husband*
Bangalore city	3·6	4·1
Rural plains sample	4·7	4·6

B. Proportion of parents who say they want no more children, by present number of children:

Number of children living	*Percentage of parents who want no more*			
	Bangalore city		*Rural plains*	
	Wife	*Husband*	*Wife*	*Husband*
0	5	10	2	2
1–3	25	42	14	26
4–6	56	72	48	59

C. Number of children desired (from the point of view of finances) in several districts near Poona:[38]

Number of children desired	*Nasik*		*Kalaba*	
	Rural (%)	*Urban* (%)	*Rural* (%)	*Urban* (%)
Less than 3	20	27	9	18
4–5	38	46	35	34
6 or more	17	19	11	14
Until a son comes	2	2	2	3
Until 2 or 3 sons come	17	4	12	10
As many as are born	5	1	19	6
Could not state a number	1	1	12	15
Total	100	100	100	100

3. Jamaica[39]

A. Answers to question 'If you could live your life over, how many children would you like to have?':

Number of children	*% of women wanting this number*
3 or less	50
4	29
5 or more	18
Up to God, fate	3
Total	100

B. Percentage of women wanting more children, by number of children they have now:

Number of living children	*% wanting more children*
0	92
1	64
2	44
3	32
4 or 5	20
6 or more	16

4. Taichung, Taiwan[40]

A. Answers to question 'How many children would you like if you could (start over) and have just the number you want?':

Number of children wanted	*% wanting this number*	
	Wife	*Husband*
less than 2	0	0
2	9	7
3	33	32
4	56	50
5	12	12
6 or more	4	6
Up to God, fate	1	*
Cannot say	2	4
Total	117	111

*Less than 1 per cent.

B. Number of children of each sex considered ideal for Taiwanese couples by wife:

Number of children	*% choosing each number for*	
	Sons	*Daughters*
0	0	0
1	2	28
2	71	63
3	20	3
4	1	0
Either sex satisfactory	2	2
Up to God, fate	1	1
Not ascertained	3	3
Total	100	100

C. Percentage of wives and husbands who have more children than they want, by present number of children:

Present number of children	*% who would prefer fewer children than they have*	
	Wife	*Husband*
1	0	0
2	0	2
3	2	4
4	12	12
5 or more	61	55

APPENDIX III. CLASSIFICATION OF THE 'INTERMEDIATE' VARIABLES AFFECTING FERTILITY

Taken from K. Davis and J. Blake, 'Social Structure and Fertility: An Analytic Framework', *Economic Development and Cultural Change*, vol. 4, no. 3 (Apr 1956) 211–35.

I. Factors affecting exposure to intercourse

A. Those governing the formation and dissolution of unions in the reproductive period.

1. Age of entry into sexual unions.
2. Permanent celibacy: proportion of women never entering sexual unions.
3. Amount of reproductive period spent after or between unions.
   (*a*) When unions are broken by divorce, separation or desertion.
   (*b*) When unions are broken by death of husband.

B. Those governing the exposure to intercourse within unions.

1. Voluntary abstinence.
2. Involuntary abstinence (from impotence, illness, unavoidable but temporary separations).
3. Coital frequency (excluding period of abstinence).

II. Factors affecting exposure to conception

A. Fecundity or infecundity, as affected by involuntary causes.

B. Use or non-use of contraception.
   1. By mechanical and chemical means.
   2. By other means.

C. Fecundity or infecundity, as affected by voluntary causes (sterilisation, subincision, medical treatment, etc.).

III. Factors affecting gestation and successful parturition

A. Foetal mortality from voluntary causes.

B. Foetal mortality from involuntary causes.

APPENDIX IV. SELECTED RESULTS OF A STUDY OF FACTORS ASSOCIATED WITH APPROVAL AND USE OF FAMILY PLANNING IN TAICHUNG, TAIWAN (PILOT STUDY OF 241 COUPLES WITH WIFE 25 TO 29 YEARS OLD)

Characteristics of wife or of couple	*% of couples who have practised some form of family planning*	*% of couples who approve of family planning unconditionally*
Wife's education:		
none	18	56
some primary	20	47
primary graduate	26	79
some secondary school	60	74
secondary school graduate	71	83
Frequency of reading daily newspaper – wife:		
never	19	61
sometimes	33	92
once or twice a week	41	86
daily	58	75
Family expenditures per month (N.T.):		
less than 1000	26	64
1000–1999	34	75
2000–2999	41	67
3000 or more	37	63
Farm background – wife:		
some	31	67
none	40	73
No. of 'modern' consumer objects owned:		
less than 4	18	50
4	17	77
5	21	76
6	41	69
7	47	77
8 or 9	48	66
Type of family:		
joint	27	60
stem	27	68
nuclear	40	72

1. T. E. Smith, 'The Cocos–Keeling Islands: A Demographic Laboratory', *Population Studies*, vol. 14 (1960).

2. For example, birth-rates of sixty or more are reported for several African populations in the United Nations, *Demographic Yearbook* (1961), on the basis of sample surveys.

3. *Royal Commission on Population Report* (London: H.M.S.O., 1949) p. 24. Women recorded as married in the fertility census of 1911 and born in the period 1841–45 reported an average of 5·71 live births.

4. For example, in the United Nations, *The Mysore Population Study* (New York: The United Nations, 1962; and in N. V. Sovani and K. Dandekar, *Fertility Survey of Nasik, Kolaba, and Satara Districts* (Poona: Gokhale Institute of Politics and Economics, Publication no. 31, 1955).

5. Cf. J. Eaton and A. Mayer, *Man's Capacity to Reproduce* (The Free Press, Glencoe, Ill., 1954) pp. 20 ff.

6. For example, K. Dandekar, *Demographic Survey of Six Rural Communities* (Poona: Gokhale Institute of Politics and Economics, 1959); P. K. Hatt, *Backgrounds of Human Fertility in Puerto Rico: A Sociological Survey* (Princeton: Princeton U.P., 1952); R. Hill, J. M. Stycos and K. W. Back, *The Family and Population Control: a Puerto Rican Experiment in Social Change* (Chapel Hill: University of North Carolina Press, 1959); S. B. Mukherjee, *Studies on Fertility Rates in Calcutta* (Calcutta: Bookland Private Ltd, 1961); B. Singh, *Five Years of Family Planning in the Country-side* (Lucknow: J. K. Institute of Sociology and Human Relations, 1958); J. M. Stycos and K. Back, *Prospects for Fertility Reduction* (New York: The Conservation Foundation, 1957); L. Tabah and R. Samuel, 'Preliminary Findings of a Survey of Fertility and Attitudes Toward Family Formation in Santiago', in Milbank Memorial Fund, *Research in Family Planning* (Princeton: Princeton U.P., 1962); United Nations: *The Mysore Population Study* (New York: United Nations, 1962); D. Yaukey, *Fertility Differences in a Modernising Country* (Princeton: Princeton U.P. 1961).

A more complete annotated bibliography of such studies appears in *The Society of Human Fertility: A Trend Report and Bibliography*, vols. 10/11, no. 2 (1961–2), in the series *Current Sociology* published by the International Sociological Association with the support of UNESCO.

7. A. J. Richards and P. Reining, 'Reports on Fertility Surveys in Buganda and Buhaya, 1952', pp. 351–404. In F. Lorimer *et al.*, *Culture and Human Fertility* (Paris: UNESCO, 1954).

8. C. S. Ford, *A Comparative Study of Human Reproduction* (New Haven: Yale University Publications in Anthropology, no. 32).

9. K. Davis and J. Blake, 'Social Structure and Fertility: An Analytic Framework', *Economic Development and Cultural Change*, vol. 4, no. 3 (April 1956) 211–35.

10. A. J. Coale and C. Y. Tye: 'The Significance of Age Patterns of Fertility in High Fertility Populations', *Milbank Memorial Fund Quarterly*, vol. 39, no. 4 (October 1961) 631–46.

11. For example, D. E. C. Eversley, 'Population and Economic Growth in England before the "takeoff".' In *Contributions and Communications to the First International Conference of Economic History* (Stockholm, 1960) pp. 457–73; H. J. Habakkuk, 'The Economic History of Modern Britain', *J. Econ. Hist.* vol. 18, no. 4 (December 1958) 486–501; J. T. Krause, 'Some Implications of Recent Work in Historical Demography', vol. 1, *Comp. Stud. Soc. Hist.*, no. 2 (January 1959) 164–88.

12. A. M. Carr-Saunders, *The Population Problem* (Oxford: Clarendon Press, 1922). The frame of reference of this important work has influenced greatly the ideas presented in the present paper.

13. N. E. Himes, *Medical History of Contraception* (Baltimore: The Williams & Wilkins Co., 1936).

14. Op. cit. (see note 8).

15. G. Devereux, *A Study of Abortion in Primitive Societies* (New York: Julian Press, 1955).

16. For relevant evidence see H. Bergues *et al.*, *La Prevention des Naissances dans la Famille* (Paris: Presses Universitaries de France, 1960); and E. Lewis-Faning, *Report on an Inquiry into Family Limitation and Its Influence on Human Fertility During the Past Fifty Years*. Papers of the Royal Commission on Population, vol. 1 (London: H.M.S.O. 1949).

17. R. M. Pierce and G. Rowntree, 'Birth Control in Britain, pt 2,' *Population Studies*, vol. 15, no. 2 (November 1961) 121–60.

18. Op. cit. (see note 6).

19. For a general discussion of this issue see J. M. Stycos, 'A Critique of the Traditional Planned Parenthood Approach in Underdeveloped Areas', in C. V. Kiser (ed.), *Research in Family Planning* (Princeton U.P., 1962). References to Taiwanese data in this paper are all drawn from unpublished work of the Taiwan Population Studies Center in which participating personnel include Dr C. H. Yen, Dr. J. Y. Peng, Dr. Y. Takeshita, Mr. T. H. Sun and Mr. S. Y. Soong.

20. United Nations, *The Mysore Population Study*, op. cit. (see note 4).

21. F. Lorimer *et al.*, *Culture and Human Fertility* (Paris: UNESCO, 1954).

22. Op. cit. (see note 9).

23. For example, in the experimental study in Singur, India, described by K. K. Mathen, in 'Preliminary Lessons Learned from the Rural Population Control Study of Singur', in C. V. Kiser, op. cit., pp. 33–50 (see note 19).

24. The Population Problems Research Council, *Fifth Public Opinion Survey on Birth Control in Japan and Sixth Opinion Survey on Family Planning and Birth Control* (Tokyo: The Mainichi Newspapers, 1959 and 1962).

25. France and Spain are among possible exceptions.

26. D. V. Glass, 'Population Growth, Fertility and Population Policy', *The Advancement of Science* (November 1960), pp. 1–11.

27. The African rates are based on estimates made by the Princeton African Demography Project, from Census Inquiry data. They were kindly supplied by Professor W. Brass of the University of Aberdeen. The data supplied were in the form of total fertility rates. These were converted to gross reproduction rates for this presentation on the assumption of a uniform sex-ratio at birth of 106.

28. From N. K. Sarkar, 'Population Trends and Population Policy in Ceylon', *Population Studies*, vol. 9, no. 3 (March 1956) 195–216.

29. T. E. Smith, op. cit. (see note 1).

30. Cf. A. J. Coale and E. M. Hoover, *Population Growth and Economic*

Development in Low Income Countries (Princeton: Princeton U.P., 1958) p. 352.

31. Estimated by T. E. Smith, in *Population Growth in Malaya* (London: Royal Institute of International Affairs, 1952), ch. 2.

32. From unpublished calculations of the Taiwan Population Studies Center.

33. From G. W. Roberts, 'Some Aspects of Mating and Fertility in the West Indies', *Population Studies*, vol. 8, no. 3 (March 1955) 199–227. Mr Roberts shows that joint reproduction rates based on both male and female rates may be more useful than the conventional female gross reproduction rates, but only the female rates are cited here, for comparability with those of other countries.

34. From J. Eaton and A. Mayer, op. cit., p. 40 (see note 5).

35. The rates for England and Wales are from reports of the Registrar-General for England and Wales. All other figures in this set are from *Population Index* (April 1961), vol. 28, no. 2.

36. From R. Hill, J. M. Stycos, and K. W. Back, op. cit., 1959, p. 72. (see note 6).

37. United Nations, *The Mysore Population Study*, op. cit., p. 142 (see note 4).

38. W. V. Sovani and K. Dandekar, op. cit., p. 104 (see note 4).

39. J. M. Stycos and K. Back, op. cit. (see note 6).

40. From unpublished data collected by Taiwan Population Studies Center, associated with the Provincial Maternal and Child Health Institute, Taiwan.

URBANISATION AND POLITICAL PROTEST

Myron Weiner

One feature of virtually all modernising nations is the growth of political demands as mass participation increases. The modernisation process, by increasing education, mass communications and urbanisation, creates conditions for mass political participation. Indeed, it is now commonly argued by most students of modernisation that political protest, violence and extremist political behaviour typically accompany modernisation. Discontent is often great not because a society is stagnant, but because it is changing.

One of the most dramatic changes in developing areas is the growth of urban centres as migrants move from the rural countryside to the cities. And a concomitant of urban growth is an increase in violence and political extremism. *Since violence is characteristic of urban areas in many developing countries and urban areas are growing as a result of the rural influx, it is logically assumed that it is the migrants rather than 'native-born' city-dwellers who are responsible for urban political extremism and violence.* Moreover, given the typical conditions under which migrants live in urban areas, this hypothesis easily gains credibility even in the absence of any empirical verification.

A typical and eloquent statement of this hypothesis is given by the well-known English writer, Barbara Ward:

> All over the world, often long in advance of effective industrialisation, the unskilled poor are streaming away from subsistence agriculture to exchange the squalor of rural poverty for the even deeper miseries of the shanty-towns, *favellas*, and *bidonvilles* that, year by year, grow inexorably on the fringes of the developing cities. They . . . are the core of local despair and disaffection – filling the *Jeunesse* movements of the Congo, swelling the urban mobs of Rio, voting Communist in the ghastly alleys of Calcutta, everywhere undermining the all too frail structure of public order and thus retarding the economic development that can alone help their plight. Unchecked, disregarded, left to grow and fester, there is here enough explosive material to produce in the world at large the pattern of a bitter class conflict finding to an increasing degree a racial bias, erupting in guerrilla warfare, and threatening, ultimately, the security of the comfortable West.[1]

The anti-social behaviour of migrants is thus assumed, and the explanations for their behaviour are typically drawn from socialisation theory. It is widely argued, for example, that villagers live under the social controls of their tribe, caste, kinsmen and village authorities and that as villagers move into the cities, often without their family, social controls break down and the migrant is prone to commit anti-social acts, become violent and support radical protest movements.

I have attempted to test the hypothesis that protest political behaviour in urban areas is a product of rural migration by looking at the characteristics of protest voters in the city of Calcutta. Calcutta provides us with a good case study, for it is a typical example of a city made up largely of migrants where political violence and extremism are commonplace.

Like many relatively 'new' cities in the developing areas (Calcutta grew in the nineteenth century as a centre for British commerce and as the capital of India), Calcutta has a high density. It has nearly three million people, plus another two and a half in the surrounding metropolitan region. It is – by almost any criteria – a centre of political extremism. In the 1962 elections the Communists won three out of four parliamentary seats (with 52 per cent of the vote) and eight out of twenty-six state assembly seats (with 32 per cent of the vote). Several Marxist left parties (not including the Socialists) won an additional 5·1 per cent of the vote in the state assembly elections. Moreover, the city is well known throughout the subcontinent as a centre of violence; mass demonstrations against high food prices or against increments in tram fares or for wage increases are a daily occurrence and hardly a year has passed since India became independent that the city has escaped substantial political violence. In fact since the turn of the century Calcutta has been a centre for rebellious activities, first against British authority and in recent years against the Congress government. And in the most recent elections of 1967 the Communists again scored a high vote, with a rather substantial number of voters turning to the pro-Peking section of the Communist movement.

Since the Congress Party was, until the 1967 elections, the governing party in the state of West Bengal (in which Calcutta is located), and since Congress is widely seen as an ideologically moderate party (and in Calcutta as a conservative party), we have analysed the vote against Congress as one measure of political discontent. And since the bulk of the opposition vote has been largely for the Communists and Marxist left parties, it is not unreasonable to describe the anti-Congress vote as a measure of political protest, if not political radicalism. In the city as a whole the vote for the Congress candidates for the state legislative assembly was 39·7 per cent in the election of 1952, 42·6 per cent in the election of 1957 and 47·2 per cent in the 1962 election.

In the absence of survey data which could specify how individuals actually voted, we have used the alternative, though less satisfying, technique of ecological correlation for determining who voted for and who voted against the Congress Party.[2] We shall use these correlations to ascertain whether the anti-Congress vote came primarily from rural migrants to the city or from those born in the city.

Two-thirds of the population of Calcutta were born outside the city. These migrants fall into two distinct groups with quite different characteristics: (1) refugees who, shortly before or after partition, fled from East Pakistan because of communal disturbances, and (2) ordinary migrants from other portions of the state and other Indian states. Henceforth we shall use the terms *refugees* and *migrant* to refer to each of these two groups. Following a study by Sen,[3] who proposes that those who entered the city after 1934 be classified as migrants and those who have lived in the city longer or been born in the city be classified as residents, we can divide the population of Calcutta as follows:

	%
Residents	57
Refugees	17
Migrants	26
	100

HOW REFUGEES VOTE

When India was partitioned in 1947 into two successor states – Pakistan and India – the state of Bengal was partitioned as well. The larger portion, predominantly Muslim, went to Pakistan and the smaller, predominantly Hindu portion including Calcutta, went to India. Hindu minorities in Pakistan soon flocked to India, many of them to Calcutta, while Muslim minorities fled to Pakistan.

Since the census reports list refugees by ward, it is possible to determine the percentage of refugees in each of the twenty-four state legislative assembly constituencies in Calcutta in 1952. When a rank order correlation is done for the 1952 elections – which took place soon after the population transfers – Congress vote and refugee vote correlated $r = -0{\cdot}81$.

In other words, where the Congress vote was high, there were few refugees, and where refugees were numerous the Congress vote was low. In the six constituencies which Congress lost in 1952 with less than 30 per cent of the vote, the refugee population ranged from 12 to 41 per cent. Of the seven constituencies in

which Congress won a majority, five had refugee populations of less than 10 per cent and the other two only slightly more.

We find virtually the same results in the elections after 1952. In the 1957 and 1962 elections the areas won by Congress had fewer refugees than those constituencies which they lost. In the constituencies won by Congress in both elections refugees averaged 11 per cent, and in the areas with a low Congress vote refugees constituted 23 per cent. *In only one constituency with a refugee population of more than 15 per cent (slightly less than the average of 17 per cent for the city as a whole) has Congress ever won a majority of votes.*

The refugees, though migrants, are quite different from 'ordinary migrants'. Since they were forced to leave their homes in East Pakistan, they came with their families (there is an almost even sex ratio of men to women) and, unlike ordinary migrants to urban centres, were completely cut off from their original homes. They left their homes not because they were attracted by the city's employment possibilities or urban amenities, but because they were forced to do so by communal disturbances. They were 'pushed', not 'pulled', into the city.

Compared to ordinary migrants, the refugees are better educated, more frequently from urban areas, and prefer white-collar jobs. Unemployment among them is three times that of ordinary migrants (12·2 per cent in 1957 compared to 4·3 per cent for ordinary migrants and 8·5 per cent for residents).[4] But, of course, unlike the unemployed migrant, the unemployed refugee cannot return home.

HOW MIGRANTS VOTE

Excluding the refugees, approximately one-fourth or 26 per cent of Calcutta's population came to the city after 1934. More than one-third speak Bengali and came from other portions of West Bengal. Two-fifths are Hindi speakers from the nearby states of Bihar and Uttar Pradesh. A small number come from the state of Orissa, just south of Calcutta, and some are Muslims from other portions of Bengal or neighbouring states.

The census reports unfortunately do not tell us how migrants are distributed throughout the city and thus an indirect method must be used for ascertaining which constituencies are heavily settled by migrants. It is known and proven, however, that sex ratios (male to female) are higher among migrants than among residents or refugees. Among residents there are 1480 males per 1000 females and among refugees 1220 males per 1000 females, while among migrants the sex ratio varies, depending upon where the migrants are from, from 1900 to 18,500 males per 1000 females. For the city as a whole the sex ratio is 1·75. It seems valid to conclude therefore that the higher the sex ratio in any constituency, the greater the number of migrants. If constituencies are rank ordered on the basis of sex ratio, they range from 3·05 males per female to 1·20 males per female. The rank order correlation between sex ratio and Congress assembly vote in all 27 constituencies in 1957 was $r = +0{\cdot}791$, and in 1962 $r = +0{\cdot}794$. In short, a high sex ratio in a constituency, indicating a large number of migrants, correlates well with a high Congress vote. *In fact there is not a single constituency in Calcutta with a ratio of 2·0, that is twice as many men as women, which failed to elect a Congressman in the 1962 elections, and constituencies lost by Congress typically have a sex ratio of 1·2 or 1·3, suggesting that constituencies with few migrants vote against Congress.*

Of course it is possible that some third factor can explain the findings, but none of the obvious controls we tested, including percentage of voting turn-out, literacy and number of slum dwellers in the constituency, affect our findings.[5] Nor have we been able to find any significant variations between migrants who come from rural areas of West Bengal as against migrants from other states. Our data suggest that in general migrants, wherever they come from, are more likely to vote for Congress than are either refugees from Pakistan or residents born in the city. *In short, our study of voting behaviour in Calcutta does not provide evidence to support the commonly held hypothesis that migrants to urban centres are politically more extremist than city-born residents.*[6]

SOME THEORETICAL OBSERVATIONS

A simplistic theory of material deprivation would lead us to expect just the reverse of our findings. In virtually all respects, migrants are clearly worse off than the residents. A larger number of migrants live in single-member households than do original residents. They are less educated than residents. A larger proportion engage in unskilled manual work. A larger proportion of migrant earners

(77 per cent) earn less than 100 rupees a month (about $20 before devaluation) than residents (64 per cent) and correspondingly fewer migrants are in upper income levels than residents. Finally, migrants are less adequately housed and a very substantial number actually live on the city pavements. Compared to residents, therefore, the migrants are poorly clothed, housed and educated and away from their families. The classic conditions of social disorganisation and of anomie exist. Yet they vote for the governing party.

Reference group theory may provide us with a somewhat more satisfactory explanation for the behaviour of the migrants. The demands of the migrant for urban services – for good housing, adequate water supply and educational facilities – are less than those of residents. Since deprivation is a comparative matter, we must remember that the migrant is likely to find many urban services – as bad as they objectively are – an improvement over what he has had in the countryside. And so long as his reference group is that of his family and friends in his native village, then his standards may continue for some time to be those of a villager. In short, to feel a sense of deprivation, the migrant needs to be socialised into urban life, and if his reference group continues to remain a non-urban group, his discontent is likely to be less. It should also be noted that objectively his need for adequate amenities (such as education for children) are less than for the resident urban dweller since he has often come without his family. Then too the unemployment rate among migrants is lower than for residents (5·9 per cent as compared to 8·5 per cent) and, as we have noted earlier, the dissatisfied migrant – dissatisfied because he cannot find a job – need not vote against the government since he has the option of returning home to his village. Finally, the migrant often comes from rural areas where he has had more contact with Congress Party workers than with opposition party workers, and in so far as he voted Congress at home, he is likely to continue to do so in the city. The Congress Party organisation in Calcutta has been particularly active in organising and servicing migrants, and this too should be counted as a factor in reducing or at least controlling and channelling discontent among migrants.

In short, there is neither a logical nor empirical reason to assume that migrants to cities are more violent or more prone to support extremist political groups or to oppose the government than are settled residents and many reasons to expect just the reverse. To be discontented with the city, one often has to be integrated into it. Moreover, the agency which socialises the rural migrant to city life – the local political party, for example, or some other political or social welfare organisation may reduce the violence and political disorganisation which we associate with political extremism. Perhaps the only sure conclusion that our data allow is that urban life does not automatically breed discontent among rural migrants.

To this conclusion we must add one caveat. It would be misleading to conclude from these findings that nothing in the rapid influx of rural migrants to an urban area need make the urban area more violent or more a centre of extremist political behaviour. For one thing, as the flow of migrants increases, typically there is a deterioration of services and a decrease in job opportunities for the residents, thereby increasing the dissatisfaction of the residents. And for another, as migration increases, the absolute number of residents is similarly increasing as migrants settle into the city, bring their families, or marry local residents.

Our analysis of Calcutta supports the proposition that cities in developing areas with a high urbanisation rate are characterised by violence and extremist political movements, but refutes the notion that these characteristics are prevalent among migrants. Our analysis does suggest that dissatisfaction is most prevalent among the city's residents. Limited though it is, this finding does bring into question a wide range of hypotheses, many of which have been widely accepted. Until survey data are available, ecological correlations of the type used here will be the best guide for studying the political behaviour of large groups. One case does not of course prove an hypothesis, nor disprove it either, but it does suggest that until we have comparative studies of other cities we ought to suspend our supposition that migrant populations are politically extreme.

Notes

1. Barbara Ward, 'The Uses of Prosperity', *Saturday Review* (29 August 1964) pp. 191–2. A similar view is taken by one of the world's leading demographers, Philip M. Hauser, in *Urbanisation in Asia and the Far East* (Calcutta: UNESCO, 1957) p. 249, who argues that 'The very patterns which characterise the life of the newly arrived migrant in the urban area are those which are generally associated with the causes of social breakdown, including criminal behaviour'.

2. The difficulties inherent in ecological correlations are discussed by W. S. Robinson 'Ecological Correlations and the Behavior of Individuals', *American Sociological Review*, vol. 15 (June 1960). Robinson argues that it is fallacious to assume that associations computed from group means or group proportions are valid estimates of the associations that would be obtained from individual data. For a defence of Durkheim's use of ecological correlations, and by implication of the method itself, see Hanon C. Selvin, Durkheim's *Suicide: Further Thoughts on a Methodological Classic*, in Robert A. Nisbet, *Émile Durkheim* (Englewood Cliffs, New Jersey; Prentice Hall, 1965) pp. 113–36.

3. S. N. Sen, *The City of Calcutta: A Socio-Economic Survey, 1954–55* to *1957–58* (Calcutta: Brookland Private Ltd, 1960).

4. Ibid., p. 228. When refugees do hold jobs they tend to receive higher salaries than do ordinary workers, suggesting that unemployment is less among migrants because they are more willing to accept lower-status jobs. According to the survey by Sen, only 29 per cent of the refugees held unskilled jobs before migrating to Calcutta as against 82 per cent of the ordinary migrants. See Sen, ibid., p. 227.

5. It is of course possible that single men prefer Congress candidates, while married men vote for the opposition candidates, but it should be noted that single men of voting age are more likely to be migrants than are the residents.

6. Apart from voting data we have other information – unfortunately, not proof – that migrants are less likely to be political extremists than residents. According to the Police Commissioner of Calcutta, persons arrested for political crimes and rowdy activities in Calcutta are primarily resident Bengalis rather than migrants. See Myron Weiner, *The Politics of Scarcity* (Chicago: University of Chicago Press, 1962) p. 192.

Our final items by B. Berelson on the one hand and E. Katz on the other serve to illustrate the dilemma of the social scientist in the face of such a complex and yet pressing range of problems as those associated with 'the population explosion'. On the one hand, he is naturally inclined towards trying to understand the situation in all its variety. On the other hand, he knows that, as Berelson points out, 'doing academic studies on family structure, or community power structure, or innovative personalities, or communication networks . . . is not going to pay off with results in time to be useful' (*editor's emphasis*). *What he can do, however, and in effect this is what Berelson is doing, is to apply in a rough and ready way the insights gained by social scientists in other areas to the particular problems brought about by the 'population explosion'. The article by Katz, on 'The Two-step Flow of Communication', is reproduced here deliberately to drive home this point. In it Katz discusses a number of studies concerned with decision-making and the adoption of innovations among farmers, doctors, shoppers and voters. None of these concerns the mass adoption of birth control – one of the major policy decisions taken by many governments in the underdeveloped countries in recent years. Yet the understanding gained from these studies influences the way we approach the implementation of a birth control programme.*

THE TWO-STEP FLOW OF COMMUNICATION: AN UP-TO-DATE REPORT ON AN HYPOTHESIS

Elihu Katz

ANALYSIS of the process of decision-making during the course of an election campaign led the authors of *The People's Choice* to suggest that the flow of mass communications may be less direct than was commonly supposed. It may be, they proposed, that influences stemming from the mass media first reach 'opinion leaders' who, in turn, pass on what they read and hear to those of their everyday associates for whom they are influential. This hypothesis was called 'the two-step flow of communication'.[1]

The hypothesis aroused considerable interest. The authors themselves were intrigued by its implications for democratic society. It was a healthy sign, they felt, that people were still most successfully persuaded by give-and-take with other people and that the influence of the mass media was less automatic and less potent than had been assumed. For social theory, and for the design of communications research, the hypothesis suggested that the image of modern urban society needed revision. The image of the audience as a mass of disconnected individuals hooked up to the media but not to each other could not be reconciled with the idea of a two-step flow of communication implying, as it did, networks of interconnected individuals through which mass communications are channelled.

Of all the ideas in *The People's Choice*, however, the two-step flow hypothesis is probably the one that was least well documented by empirical data. And the reason for this is clear: the design of the study did not anticipate the importance which interpersonal relations would assume in the analysis of the data. Given the image of the atomised audience which characterised so much of mass media research, the surprising thing is that interpersonal influence attracted the attention of the researchers at all.[2]

In the almost seventeen years since the voting study was undertaken, several studies at the Bureau of Applied Social Research of Columbia University have attempted to examine the hypothesis and to build upon it. Four such studies will be singled out for review. These are Merton's study[3] of interpersonal influence and communications behaviour in Rovere; the Decatur study of decision-making in marketing, fashions, movie-going and public affairs, reported by Katz and Lazarsfeld;[4] the Elmira study of the 1948 election campaign reported by Berelson, Lazarsfeld and McPhee;[5] and finally, a very recent study by Katz and Menzel on the diffusion of a new drug among doctors.

These studies will serve as a framework within which an attempt will be made to report on the present state of the two-step flow hypothesis, to examine the extent to which it has found confirmation and the ways in which it has been extended, contracted

and reformulated. More than that, the studies will be drawn upon to highlight the successive strategies which have been developed in attempting to take systematic account of interpersonal relations in the design of communications research, aiming ultimately at a sort of 'survey sociometry'. Finally, these studies, plus others which will be referred to in passing, will provide an unusual opportunity to reflect upon problems in the continuity of social research.[6]

FINDINGS OF *THE PEOPLE'S CHOICE*

The starting point for this review must be an examination of the evidence in the 1940 voting study which led to the original formulation of the hypothesis. Essentially, three distinct sets of findings seem to have been involved. The first had to do with *the impact of personal influence*. It is reported that people who made up their minds late in the campaign, and those who changed their minds during the course of the campaign, were more likely than other people to mention personal influence as having figured in their decisions. The political pressure brought to bear by everyday groups such as family and friends is illustrated by reference to the political homogeneity which characterises such groups. What's more, on an average day, a greater number of people reported participating in discussion of the election than hearing a campaign speech or reading a newspaper editorial. From all of this, the authors conclude that personal contacts appear to have been both more frequent and more effective than the mass media in influencing voting decisions.[1]

The second ingredient that went into the formulation of the hypothesis concerned *the flow of personal influence*. Given the apparent importance of interpersonal influence, the obvious next step was to ask whether some people were more important than others in the transmission of influence. The study sought to single out the 'opinion leaders' by two questions: 'Have you recently tried to convince anyone of your political ideas?', and 'Has anyone recently asked you for your advice on a political question?' Comparing the opinion leaders with others, they found the opinion leaders more interested in the election. And from the almost even distribution of opinion leaders throughout every class and occupation, as well as the frequent mention by decision-makers of the influence of friends, co-workers and relatives, it was concluded that opinion leaders are to be found on every level of society and presumably, therefore, are very much like the people whom they influence.[1]

A further comparison of leaders and others with respect to mass media habits provides the third ingredient: *the opinion leaders and the mass media*. Compared with the rest of the population, opinion leaders were found to be considerably more exposed to the radio, to the newspapers and to magazines, that is, to the formal media of communication.

Now the argument is clear: If word-of-mouth is so important, and if word-of-mouth specialists are widely dispersed, and if these specialists are more exposed to the media than the people whom they influence, then perhaps 'ideas often flow from radio and print to opinion leaders and from these to the less active sections of the population'.[1]

DESIGN OF THE VOTING STUDY

For studing the flow of influence as it impinges on the making of decisions, the study design of *The People's Choice* had several advantages. Most important was the panel method which made it possible to locate changes almost as soon as they occurred and then to correlate change with the influences reaching the decision-maker. Secondly, the unit of effect, the decision, was a tangible indicator of change which could readily be recorded. But for studying that part of the flow of influence which had to do with contacts among people, the study design fell short, since it called for a random sample of individuals abstracted from their social environments. It is this traditional element in the design of survey research which explains the leap that had to be made from the available data to the hypothesis of the two-step flow of communication.

Because every man in a random sample can speak only for himself, opinion leaders in the 1940 voting study had to be located by self-designation, that is, on the basis of their own answers to the two advice-giving questions cited above.[7] In effect, respondents were simply asked to report whether or not they were opinion leaders. Much more important than the obvious problem of validity posed by this technique is the fact that it does not permit a comparison of leaders with their respective followers, but only

of leaders and non-leaders in general. The data, in other words, consist only of two statistical groupings: people who said they were advice-givers and those who did not. Therefore, the fact that leaders were more interested in the election than non-leaders cannot be taken to mean that influence flows from more interested persons to less interested ones. To state the problem drastically, it may even be that the leaders influence only each other, while the uninterested non-leaders stand outside the influence market altogether. Nevertheless, the temptation to assume that the non-leaders are the followers of the leaders is very great, and while *The People's Choice* is quite careful about this, it cannot help but succumb.[8] Thus, from the fact that the opinion leaders were more exposed to the mass media than the non-leaders came the suggestion of the two-step flow of communication; yet, manifestly, it can be true only if the non-leaders are, in fact, followers of the leaders.

The authors themselves point out that a far better method would have been based on 'asking people to whom they turn for advice on the issue at hand and then investigating the interaction between advisers and advisees. But that procedure would be extremely difficult, if not impossible, since few of the related 'leaders' and 'followers' would happen to be included in the sample.'[1] As will be shown immediately, this is perhaps the most important problem which succeeding studies have attempted to solve.

DESIGNS OF THREE SUBSEQUENT STUDIES

To this point, two aspects of the original statement of the two-step flow hypothesis have been reviewed. First of all, the hypothesis has been shown to have three distinct components, concerning respectively the impact of personal influence; the flow of personal influence; and the relationship of opinion leaders to the mass media. The evidence underlying each has been examined. Secondly, the design of the study has been recalled in order to point up the difficulty that arises from attempting to cope with the fundamentally new problem of incorporating *both* partners to an influence transaction into a cross-sectional study.

From this point forward, the major focus will turn to those studies that have succeeded *The People's Choice*. We will first report the different ways in which three of the four studies selected for review approached the problem of designing research on interpersonal influence.[9] Thereafter, the substantive findings of the several studies will be reviewed and evaluated so as to constitute an up-to-date report on the accumulating evidence for and against the hypothesis of the two-step flow of communication.

1. The Rovere study

Undertaken just as the 1940 voting study was being completed, the earliest of the three studies was conducted in a small town in New Jersey. It began by asking a sample of 86 respondents to name the people to whom they turned for information and advice regarding a variety of matters. Hundreds of names were mentioned in response, and those who were designated four times or more were considered opinion leaders. These influentials were then sought out and interviewed.[3]

Here, then, is the initial attempt, on a pilot scale, to solve the problem or research design posed by *The People's Choice*. To locate influentials, this study suggests, begin by asking somebody 'Who influences you?' and proceed from the persons influenced to those who are designated as influential.

Two important differences between this study and the 1940 voting study must be pointed out. First, there is a difference in the conception of opinion leadership. Whereas the voting study regards any advice-giver as an opinion leader if he influences even one other person (such as a husband telling his wife for whom to vote), the leaders singled out by the criterion employed in Rovere were almost certainly wielders of wider influence.

Secondly, the voting study, at least by implication, was interested in such questions as the extent of the role of interpersonal influence in decision-making and its relative effectiveness compared to the mass media. The Rovere study took for granted the importance of this kind of influence, and proceeded to try to find people who play key roles in its transmission.

A final point to make in connection with the design of this study is that it makes use of the initial interviews almost exclusively to *locate* opinion leaders and hardly at all to explore the *relationships* between leaders and followers. Once the leaders were designated, almost exclusive attention was given to classifying them into different types, studying the communications behaviour of the different types and the interaction among the leaders themselves,

but very little attention was given to the interaction between the leaders and the original informants who designated them.

2. The Decatur study

The Decatur study, carried out in 1945–6, tried to go a step farther.[4] Like the voting study, but unlike Rovere, it tried to account for decisions – specific instances in which the effect of various influences could be discerned and assessed. Like Rovere, but unlike the voting study, it provided for interviews with the persons whom individuals in the initial sample had credited as influential in the making of recent decisions (in the realms of marketing, movie-going and public affairs). The focus of the study this time was not on the opinion leaders alone, but (1) on the relative importance of personal influence and (2) on the person who named the leader as well as the leader – the adviser–advisee dyad.

Ideally, then, this study could ask whether opinion leaders tended to be from the same social class as their followers or whether the tendency was for influence to flow from the upper classes downwards. Were members of the dyads likely to be of the same age, the same sex, etc.? Was the leader more interested in the particular sphere of influence than his advisee? Was he more likely to be exposed to the mass media?

Just as the dyad could be constructed by proceeding from an advisee to his adviser, it was also possible to begin the other way around by talking first to a person who claimed to have acted as an adviser, and then locating the person he said he had influenced. The Decatur study tried this too. Using the same kind of self-designating questions employed in the voting study, persons who designated themselves as influential were asked to indicate the names of those whom they had influenced. By 'snow-balling' to the people thus designated, there arose the opportunity not only to study the interaction between adviser and advisee but also to explore the extent to which people who designated themselves as influential were confirmed in their self-evaluations by those whom they allegedly had influenced. Proceeding in this way, the researchers hoped to be able to say something about the validity of the self-designating technique.[4,10]

The authors of *The People's Choice* had said that 'asking people to whom they turn and then investigating the interaction between advisers and advisees . . . would be extremely difficult if not impossible'. And, in fact, it proved to be extremely difficult. Many problems were encountered in the field work, the result of which was that not all the 'snowball' interviews could be completed.[11] In many parts of the analysis of the data, therefore, it was necessary to revert to comparisons of leaders and non-leaders, imputing greater influence to groups with higher concentrations of self-designated leadership. Yet, in principle, it was demonstrated that a study design taking account of interpersonal relations was both possible and profitable to execute.

But about the time it became evident that this goal was within reach, the goal itself began to change. It began to seem desirable to take account of chains of influence longer than those involved in the dyad; and hence to view the adviser–advisee dyad as one component of a more elaborately structured social group.

These changes came about gradually and for a variety of reasons. First of all, findings from the Decatur study and from the later Elmira study revealed that the opinion leaders themselves often reported that their own decisions were influenced by still other people.[5] It began to seem desirable, therefore, to think in terms of the opinion leaders of opinion leaders.[4] Secondly, it became clear that opinion leadership could not be viewed as a trait which some people possess and others do not, although the voting study sometimes implied this view. Instead, it seemed quite apparent that the opinion leader is influential at certain times and with respect to certain substantive areas by virtue of the fact that he is 'empowered' to be so by other members of his group. Why certain people are chosen must be accounted for not only in demographic terms (social status, sex, age, etc.) but also in terms of the structure and values of the groups of which both adviser and advisee are members. Thus, the unexpected rise of young men to opinion leadership in traditional groups, when these groups faced the new situations of urbanisation and industrialisation, can be understood only against the background of old and new patterns of social relations within the group and of old and new patterns of orientation to the world outside the group.[12] Reviewing the literature of small group research hastened the formulation of this conception.[4]

One other factor shaped the direction of the new programme as well. Reflecting upon the Decatur study, it became clear that while one could talk about the role of various influences in the making of fashion *decisions by individuals*, the study design was not

adequate for the study of fashion in the aggregate – *fashion as a process of diffusion* – as long as it did not take account of either the content of the decision or the time factor involved. The decisions of the 'fashion changers' studied in Decatur might have cancelled each other out: while Mrs X reported a change from Fashion A to Fashion B, Mrs Y might have been reporting a change from B to A. What is true for fashion is true for any other diffusion phenomenon: to study it, one must trace the flow of some specific item over time. Combining this interest in diffusion with that of studying the role of more elaborate social networks of communication gave birth to a new study which focused on (1) a specific item, (2) diffusion over time, (3) through the social structure of an entire community.

3. The drug study

This study was conducted to determine the way in which doctors make decisions to adopt new drugs. This time, when it came to designing a study which would take account of the possible role of interpersonal influence among physicians, it became clear that there were so few physicians (less than one and one-half per 1000 population) that it was feasible to interview all members of the medical profession in several cities. If all doctors (or all doctors in specialties concerned with the issue at hand) could be interviewed, then there would be no doubt that all adviser–advisee pairs would fall within the sample. All such pairs could then be located within the context of larger social groupings of doctors, which could be measured by sociometric methods.

Doctors in the relevant specialties in four midwestern cities were interviewed. In addition to questions on background, attitudes, drug-use, exposure to various sources of information and influence and the like, each doctor was also asked to name the three colleagues he saw most often socially, the three colleagues with whom he talked most frequently about cases and the three colleagues to whom he looked for information and advice.[13]

In addition to the opportunity of mapping the networks of interpersonal relations, the drug study also provided for the two other factors necessary for a true diffusion study: attention to a specific item in the course of gaining acceptance, and a record of this diffusion over time. This was accomplished by means of an audit of prescriptions on file in the local pharmacies of the cities studied, which made it possible to date each doctor's earliest use of a particular new drug – a drug which had gained widespread acceptance a few months before the study had begun. Each doctor could thus be classified in terms of the promptness of his decision to respond to the innovation, and in terms of other information provided by the prescription audit.

Altogether, compared with the earlier studies, the drug study imposes a more objective framework – both psychological and sociological – on the decision. First of all, the decision-maker himself is not the only source of information concerning his decision. Objective data from the prescription record are used as well. Secondly, the role of different influences is assessed not only on the basis of the decision-maker's own reconstruction of the event, but also on the basis of objective correlations from which inferences concerning the flow of influence can be drawn. For example, doctors who adopted the new drug early were more likely to be participants in out-of-town medical specialty meetings than those who adopted it later.

Similarly, it is possible to infer the role of social relations in doctor's decision-making not only from the doctor's own testimony concerning the role of social influences but also from the doctor's 'location' in the interpersonal networks mapped by the sociometric questions. Thus, on the basis of sociometric data, it is possible to classify doctors according to their integration into the medical community, or the degree of their influence, as measured by *the number of times* they are named by their colleagues as friends, discussion partners and consultants. They can also be classified according to their membership in one or another network or clique, as indicated by *who* names them. Using the first measure makes it possible to investigate whether or not the more influential doctors adopt a drug earlier than those who are less influential. From the second kind of analysis one can learn, for example, whether or not those doctors who belong to the same sub-groups have similar drug-use patterns. In this way, it becomes possible to weave back and forth between the doctor's own testimony about his decisions and the influences involved, on the one hand, and the more objective record of his decisions and of the influences to which he has been exposed, on the other hand.

Note that the networks of social relations in this study are mapped 'prior' to the introduction of the new drug being studied,

in the sense that friendship, consultation and so on are recorded independently of any particular decision the doctor has made. The study is concerned with the potential relevance of various parts of these sociometric structures to the transmission of influence. For example, it is possible to point to the parts of the structure which are 'activated' upon the introduction of a new drug, and to describe the sequence of diffusion of the drug as it gains acceptance by individuals and groups in the community. While the Decatur study could hope to examine only the particular face-to-face relationship which had been influential in a given decision, the drug study can locate this relationship against the background of the entire web of *potentially* relevant relationships within which the doctor is embedded.

THE FINDINGS OF STUDIES SUBSEQUENT TO *THE PEOPLE'S CHOICE*

Having examined the *designs* of these studies, the next step is to explore their *findings* in so far as these are relevant to the hypothesis about the two-step flow of communication. It will be useful to return to the three categories already singled out in discussing *The People's Choice*: (1) the impact of personal influence; (2) the flow of personal influence; and (3) opinion leaders and the mass media. Evidence from the three studies just reported, as well as from the 1948 Elmira study[5] and from others, will be brought together here; but in every case the characteristics of each study's design must be borne in mind in evaluating the evidence presented.

(*a*) THE IMPACT OF PERSONAL INFLUENCE

1. Personal and the mass media influence

The 1940 study indicated that personal influence affected voting decisions more than the mass media did, particularly in the case of those who changed their minds during the course of the campaign. The Decatur study went on to explore the relative impact of personal influences and the mass media in three other realms: marketing, fashions and movie-going. Basing its conclusions on the testimony of the decision-makers themselves, and using an instrument for evaluating the relative effectiveness of the various media which entered into the decisions, the Decatur study again found that personal influence figured both more frequently and more effectively than any of the mass media.[4]

In the analysis to date, the drug study has not approached the problem of the relative effectiveness of the various media from the point of view of the doctor's own reconstruction of what went into the making of his decision. Comparing mere frequency of mention of different media, it is clear that colleagues are by no means the most frequently mentioned source. Nevertheless, exploration of the factors related to whether the doctor's decision to adopt the drug came early or late indicates that the factor most strongly associated with the time of adoption of the new drug is the extent of the doctor's integration in the medical community. That is, the more frequently a doctor is named by his colleagues as a friend or a discussion partner, the more likely he is to be an innovator with respect to the new drug. Extent of integration proves to be a more important factor than any background factor (such as age, medical school or income of patients), or any other source of influence (such as readership of medical journals) that was examined.

Investigation of why integration is related to innovation suggests two central factors: (1) interpersonal communication – doctors who are integrated are more in touch and more up-to-date; and (2) social support – doctors who are integrated feel more secure when facing the risks of innovation in medicine.[14] Thus the drug study, too, provides evidence of the strong impact of personal relations – even in the making of scientific decisions.

2. Homogeneity of opinion in primary groups

The effectiveness of interpersonal influence, as it is revealed in the studies under review, is reflected in the homogeneity of opinions and actions in primary groups. The medium of primary group communication is, by definition, person-to-person. Both of the voting studies indicate the high degree of homogeneity of political opinion among members of the same families, and among co-workers and friends. The effectiveness of such primary groups in pulling potential deviates back into line is demonstrated by the fact that those who changed their vote intentions were largely people who, early in the campaign, had reported that they intended to vote differently from their family or friends.[1,5]

The drug study, too, was able to examine the extent of homogeneity in the behaviour of sociometrically related doctors, and

was able to demonstrate that there were situations where similar behaviour could be deserved. For example, it was found that, when called upon to treat the more puzzling diseases, doctors were likely to prescribe the same drug as their sociometric colleagues. The study also showed that, very early in the history of a new drug, innovating doctors who were sociometrically connected tended to adopt the new drug at virtually the same time. This phenomenon of homogeneity of opinion or behaviour among interacting individuals confronting an unclear or uncertain situation which calls for action has often been studied by sociologists and social psychologists.[15]

3. The various roles of the media

The 1940 voting study explored some of the reasons why personal influence might be expected to be more influential in changing opinions than the mass media: It is often non-purposive; it is flexible; it is trustworthy. It was suggested that the mass media more often play a reinforcing role in the strengthening of predispositions and of decisions already taken. Nevertheless, it was assumed that the various media and personal influence are essentially competitive, in the sense that a given decision is influenced by one *or* the other. The Decatur study tended towards this assumption too, but at one point the study does attempt to show that different media play different parts in the decision-making process and take patterned positions in a sequence of several influences. The drug study elaborates on the roles of the media even further, distinguishing between media that 'inform' and media that 'legitimate' decisions. Thus in doctors' decisions, professional media (including colleagues) seem to play a legitimating role, while commercial media play an informing role.

(*b*) THE FLOW OF PERSONAL INFLUENCE

The 1940 voting study found that opinion leaders were not concentrated in the upper brackets of the population but were located in almost equal proportions in every social group and stratum. This finding led to efforts in subsequent studies to establish the extent to which this was true in areas other than election campaigns and also to ascertain what it is that *does* distinguish opinion leaders from those whom they influence.

The first thing that is clear from the series of studies under review is that the subject matter concerning which influence is transmitted has a lot to do with determining who will lead and who follow. Thus, the Rovere study suggests that within the broad sphere of public affairs one set of influentials is occupied with 'local' affairs and another with 'cosmopolitan' affairs. The Decatur study suggests that in marketing, for example, there is a concentration of opinion leadership among older women with larger families, while in fashions and movie-going it is the young, unmarried girl who has a disproportionate chance of being turned to for advice. There is very little overlap of leadership: a leader in one sphere is not likely to be influential in another unrelated sphere as well.[4]

Yet, even when leadership in one or another sphere is heavily concentrated among the members of a particular group – as was the case with marketing leadership in Decatur – the evidence suggests that people still talk, most of all, to others like themselves. Thus, while the marketing leaders among the older 'large-family wives' also influenced other kinds of women, most of their influence was directed to women of their own age with equally large families. In marketing, fashions and movie-going, furthermore, there was no appreciable concentration of influentials in any of the three socio-economic levels. Only in public affairs was there a concentration of leadership in the highest status, and there was some slight evidence that influence flows from this group to individuals of lower status. The Elmira study also found opinion leaders in similar proportions on every socio-economic and occupational level and found that conversations concerning the campaign went on, typically, between people of similar age, occupation and political opinion.

What makes for the concentration of certain kinds of opinion leadership within certain groups? And when influential and influencee are outwardly alike – as they so often seem to be – what, if anything, distinguishes one from the other? Broadly, it appears that influence is related (1) to the *personification of certain values* (who one is); (2) to *competence* (what one knows); and (3) to *strategic social location* (whom one knows). Social location, in turn, divides into whom one knows within a group; and 'outside'.

Influence is often successfully transmitted because the influencee wants to be as much like the influential as possible.[16] That the

young, unmarried girls are fashion leaders can be understood easily in a culture where youth and youthfulness are supreme values. This is an example where 'who one is' counts very heavily.

But 'what one knows' is no less important.'[17] The fact is that older women, by virtue of their greater experience, are looked to as marketing advisers and that specialists in internal medicine – the most 'scientific' of the practising physicians – are the most frequently mentioned opinion leaders among the doctors. The influence of young people in the realm of movie-going can also be understood best in terms of their familiarity with the motion picture world. The Elmira study found slightly greater concentrations of opinion leadership among the more educated people on each socio-economic level, again implying the importance of competence. Finally, the influence of the 'cosmopolitans' in Rovere rested on the presumption that they had large amounts of information.

It is, however, not enough to be a person whom others want to emulate, or to be competent. One must also be accessible. Thus, the Decatur study finds gregariousness – 'whom one knows' – related to every kind of leadership. The Rovere study reports that the leadership of the 'local' influentials is based on their central location in the web of interpersonal contacts. Similarly, studies of rumour transmission have singled out those who are 'socially active' as agents of rumour.[18]

Of course, the importance of whom one knows is not simply a matter of the number of people with whom an opinion leader is in contact. It is also a question of whether the people with whom he is in touch happen to be interested in the area in which his leadership is likely to be sought. For this reason, it is quite clear that the greater interest of opinion leaders in the subjects over which they exert influence is not a sufficient explanation of their influence. While the voting studies as well as the Decatur study show leaders to be more interested, the Decatur study goes on to show that interest alone is not the determining factor.[4] In fashion, for example, a young unmarried girl is considerably more likely to be influential than a matron with an equally great interest in clothes. The reason, it is suggested, is that a girl who is interested in fashion is much more likely than a matron with an equally high interest to know other people who share her preoccupation, and thus is more likely than the matron to have followers who are interested enough to ask for her advice. In other words, it takes two to be a leader – a leader and a follower.

Finally, there is the second aspect of 'whom one knows'. An individual may be influential not only because people within his group look to him for advice but also because of whom he knows outside his group.[19] Both the Elmira and Decatur studies found that men are more likely than women to be opinion leaders in the realm of public affairs and this, it is suggested, is because they have more of a chance to get outside the home to meet people and talk politics. Similarly, the Elmira study indicated that opinion leaders belonged to more organisations, more often knew workers for the political parties and so on, than did others. The drug study found that influential doctors could be characterised in terms of such things as their more frequent attendance at out-of-town meetings and the diversity of places with which they maintained contact, particularly far-away places. It is interesting that a study of the farmer-innovators responsible for the diffusion of hybrid seed-corn in Iowa concluded that these leaders also could be characterised in terms of the relative frequency of their trips out of town.[20]

(*c*) THE OPINION LEADERS AND THE MASS MEDIA

The third aspect of the hypothesis of the two-step flow of communication states that opinion leaders are more exposed to the mass media than are those whom they influence. In *The People's Choice* this is supported by reference to the media behaviour of leaders and non-leaders.

The Decatur study corroborated this finding, and went on to explore two additional aspects of the same idea.[4] First of all, it was shown that leaders in a given sphere (fashions, public affairs, etc.) were particularly likely to be exposed to the media appropriate to that sphere. This is essentially a corroboration of the Rovere finding that those who proved influential with regard to 'cosmopolitan' matters were more likely to be readers of national news magazines, but that this was not at all the case for those influential with regard to 'local' matters. Secondly, the Decatur study shows that at least in the realm of fashions, the leaders are not only more exposed to the mass media, but are also more affected by them in their own decisions. This did not appear to be the case in other realms, where opinion leaders, though more

exposed to the media than non-leaders, nevertheless reported personal influence as the major factor in their decisions. This suggests that in some spheres considerably longer chains of person-to-person influence than the dyad may have to be traced back before one encounters any decisive influence by the mass media, even though their contributory influence may be perceived at many points. This was suggested by the Elmira study too. It found that the leaders, though more exposed to the media, also more often reported that they sought information and advice from other persons.[5]

Similarly, the drug study showed that the influential doctors were more likely to be readers of a large number of professional journals and valued them more highly than did doctors of lesser influence. But at the same time, they were as likely as other doctors to say that local colleagues were an important source of information and advice in their reaching particular decisions.

Finally, the drug study demonstrated that the more influential doctors could be characterised by their greater attention not only to medical journals, but to out-of-town meetings and contacts as well. This finding has already been discussed in the previous section treating the *strategic location* of the opinion leader with respect to 'the world outside' his group. Considering it again under the present heading suggests that the greater exposure of the opinion leader to the mass media may only be a special case of the more general proposition that opinion leaders serve to relate their groups to relevant parts of the environment through whatever media happen to be appropriate. This more general statement makes clear the similar functions of big city newspapers for the Decatur fashion leader; of national news magazines for the 'cosmopolitan' influentials of Rovere; of out-of-town medical meetings for the influential doctor; and of contact with the city for the farmer-innovator in Iowa[21] as well as for the newly risen, young opinion leaders in underdeveloped areas throughout the world.[22]

CONCLUSIONS

Despite the diversity of subject matter with which they are concerned, the studies reviewed here constitute an example of continuity and cumulation both in research design and theoretical commitment. Piecing together the findings of the latter-day studies in the light of the original statement of the two-step flow hypothesis suggests the following picture.

Opinion leaders and the people whom they influence are very much alike and typically belong to the same primary groups of family, friends and co-workers. While the opinion leader may be more interested in the particular sphere in which he is influential, it is highly unlikely that the persons influenced will be very far behind the leader in their level of interest. Influentials and influencees may exchange roles in different spheres of influence. Most spheres focus the group's attention on some related part of the world outside the group, and it is the opinion leader's function to bring the group into touch with this relevant part of its environment through whatever media are appropriate. In every case, influentials have been found to be more exposed to these points of contact with the outside world. Nevertheless, it is also true that, despite their greater exposure to the media, most opinion leaders are primarily affected not by the communication media but by still other people.

The main emphasis of the two-step flow hypothesis appears to be on only one aspect of interpersonal relations – interpersonal relations as channels of communication. But from the several studies reviewed, it is clear that these very same interpersonal relations influence the making of decisions in at least two additional ways. In addition to serving as networks of communication, interpersonal relations are also sources of pressure to conform to the group's way of thinking and acting, as well as sources of social support. The workings of group pressure are clearly evident in the homogeneity of opinion and action observed among voters and among doctors in situations of unclarity or uncertainty. The social support that comes from being integrated in the medical community may give a doctor the confidence required to carry out a resolution to adopt a new drug. Thus, interpersonal relations are (1) channels of information, (2) sources of social pressure and (3) sources of social support, and each relates interpersonal relations to decision-making in a somewhat different way.[23]

The central methodological problem in each of the studies reviewed has been how to take account of interpersonal relations and still preserve the economy and representativeness which the random, cross-sectional sample affords. Answers to this problem

range from asking individuals in the sample to describe the others with whom they interacted (Elmira), to conducting 'snowball' interviews with influential–influencee dyads (Decatur), to interviewing an entire community (drug study). Future studies will probably find themselves somewhere in between. For most studies, however, the guiding principle would seem to be to build larger or smaller social molecules around each individual atom in the sample.[24]

Notes

1. P. F. Lazarsfeld, B. R. Berelson and H. Gaudet, *The People's Choice* (New York, 1948).

2. For the discussion of the image of the atomised audience and the contravening empirical evidence, see E. Katz and P. F. Lazarsfeld, *Personal Influence: The Part Played by People in the Flow of Mass Communications* (New York, 1955); E. Friedson, 'Communications Research and the Concept of the Mass', *Amer. Sociol. Rev.*, vol. 18 (1953) 313–17; and M. Janowitz, *The Urban Press in a Community Setting* (New York, 1952).

3. R. K. Merton, 'Patterns of Influence: a Study of Interpersonal Influence and Communications Behavior in a Local Community', in P. F. Lazarsfeld and F. N. Stanton (eds), *Communications Research, 1948–1949* (New York, 1949).

4. E. Katz and P. F. Lazarsfeld (1955), op. cit.

5. B. R. Berelson, P. F. Lazarsfeld and W. N. McPhee, *Voting: A Study of Opinion Formation in a Presidential Campaign* (Chicago, 1954).

6. Other authors who have drawn upon the concepts of opinion leadership and the two-step flow of communication, and developed them further are: M. Riley and J. Riley, 'A Sociological Approach to Communications Research', *Publ. Opin. Quart.*, vol. 15 (1951) 445–60; S. N. Eisenstadt, 'Communications Processes among Immigrants in Israel', *Publ. Opin. Quart.*, vol. 16 (1952) 42–58; D. Riesman, *The Lonely Crowd* (New Haven, 1952); and L. A. Handel, *Hollywood Looks at its Audience* (Urbana, Illinois, 1950). The programme of research in international communications at the Bureau of Applied Social Research has given considerable attention to opinion leadership; see C. Y. Glock, 'The Comparative Study of Communications and Opinion Formation', *Publ. Opin. Quart.*, vol. 16 (1952–3) 512–23; J. M. Stycos 'Patterns of communication in a rural Greek village', *Publ. Opin. Quart.*, vol. 16 (1952) 59–70; and the forthcoming book by Daniel Lerner, Paul Berkman and Lucille Pevsner, *Modernizing the Middle East*. Forthcoming studies by Peter H. Rossi and by Robert D. Leigh and Martin A. Trow are also concerned with the interplay of personal and mass-media influences in local communities.

7. Strictly speaking, of course, if a respondent reports whether or not he is a leader he is not speaking for himself but for his followers, real or imagined. Furthermore, it ought to be pointed out for the record that it is sometimes possible for a respondent to speak for others besides himself. The voting studies, for example, ask respondents to report the vote-intentions of other family members, of friends, of co-workers, though this procedure is of undetermined validity.

8. There is an alternative procedure which is something of an improvement. Respondents can be asked not only whether they have given advice but whether they have taken advice. This was done in the Decatur and Elmira studies which are cited below. Thus the non-leaders can be classified in terms of whether or not they are in the influence market at all, that is, whether or not they are 'followers'.

9. The Elmira study will be omitted at this point because its design is essentially the same as that of the 1940 voting study except for the important fact that it obtained from each respondent considerable more information about the vote-intentions of others in his environment, the kinds of people he talks with, etc., than was done in *The People's Choice*.

10. About two-thirds of the alleged influences confirmed the fact that a conversation had taken place between themselves and the self-designated influential on the subject matter in question. Of these, about 80 per cent further confirmed that they had received advice. The extent of confirmation is considerably less in the realm of public affairs than it is in marketing or fashion.

11. Partly this was due to inability to locate the designated people, but partly too, to the fact that original respondents did not always know the person who had influenced them, as is obvious for example, in the case of a woman copying another woman's hat style, etc.

12. See for example, S. N. Eisenstadt 'Communication Systems and Social Structure; An Exploratory Study', *Publ. Opinion. Quart.*, vol. 19 (1955) 153–67; Glock, op. cit. The Rovere study too, takes careful account of the structure of social relations and values in which influentials are embedded, and discusses the various avenues to influentiality open to different kinds of people.

13. H. Menzel and E. Katz, 'Social Relations and Innovation in the Medical Profession', *Publ. Opin. Quart.*, vol. 19 (1955) 337–52.

14. On the relationship between social integration and self-confidence in a work situation, see P. M. Blau, *The Dynamics of Bureaucracy* (Chicago, 1955).

15. That men, faced with an unstructured situation look to each other to establish a social reality in terms of which they act, is a central theme in the work of Durkheim, Kurt Lewin and his disciples, H. S. Sullivan ('consensual validation') and in the studies of Sherif, Asch and others.

16. That leaders are, in a certain sense, the most conformist members of their groups – upholding whatever norms and values are central to the group – is a proposition which further illustrates this point, see C. P. Marsh and A. Coleman, 'Farmers' Practice Adoption Rates in Relation to Adoption Rates of Leaders', *Rural Soc.*, vol. 19 (1954) 180–3.

17. The distinction between 'what' and 'whom' one knows is used by Merton, op. cit.

18. G. W. Allport and L. J. Postman, *The Psychology of Rumor* (New York, 1943).

19. It is interesting that a number of studies have found that the most

integrated persons within a group are also likely to have more contacts outside the group than others. One might have expected the more marginal members to have more contacts outside, see Blau, op. cit.

20. B. Ryan and N. Gross, *Acceptance and Diffusion of Hybrid Seed Corn in Two Iowa Communities* (Ames, Iowa: Iowa State College of Agriculture and Mechanic Arts), Research Bulletin 372 (1942).

21. Ryan and Gross, op. cit., choose to explain 'trips to the city' as another index of the non-traditional orientation of which innovation itself is also an index. In the case of the out-of-town meetings, trips to out-of-town centres of learning, etc., the latter were also mentioned as key sources of advice by doctors who were innovators and influentials.

22. See the forthcoming book by Lerner *et al.*, op. cit.

23. These further dimensions of interpersonal relations can be further illustrated by reference to studies which represent the 'pure type' of each dimension. Studies of rumour flow illustrate the 'channels' dimension; see, for example, J. L. Moreno, *Who Shall Survive* (Beacon, New York, 1953); L. Festinger, S. Schachter and K. Back, *Social Pressures in Informal Groups* (New York, 1950); and Blau, op. cit.

24. Various ways of accomplishing this have been discussed for the past two years in a staff seminar on 'relational analysis' at the Bureau of Applied Social Research, see S. M. Lipset, M. A. Trow and J. S. Coleman, *Union Democracy* (New York, 1956), which illustrates one approach in its study of printers within the varying social contexts of the shops in which they are employed; see also Riley and Riley, op. cit.

ON FAMILY PLANNING COMMUNICATION

Bernard Berelson

I

JUST a year and a half ago, I first came into direct professional contact with the world population problem, when the Population Council asked me to see if material from the communication field could be applied to the family planning being developed around the world. All I can do is to summarise where I have come out after eighteen months of trying to do the job. For mnemonic purposes I put my tentative conclusion under ten headings, and these ten points are necessarily compressed – perhaps overcompressed and oversimplified.

Our task is to learn how the birth rate in the so-called underdeveloped areas of the world can be brought down, through voluntary contraceptive actions of the couples involved, by a substantial amount and in a short period of time. Just to indicate order of magnitude, by 'substantial amount' and 'short time' I mean something like ten points off the birth rate in five years or fifteen points off in ten years – say, move from a birth rate of 45 to 35 in five to seven years. Can that be done? And if so, how? *Whatever answer we get from our studies, that answer must be realistically applicable to the society as a whole or to large segments of it.* 'Realistically applicable' means that the recommended programme must be practicable when gauged against the resources in money and in personnel that are available for the task. So we are searching for efficient means and not simply effective ones. Finding out that you can lower the birth rate in a little corner of the society at an expenditure, in money or in personnel, of ten times the amount you would have available to do it across the society as a whole constitutes an advance but is far short of the knowlege we need.

When the problem is put this way, it is hard to think of a greater challenge to social research. What we are up against is nothing less than trying to change the behaviour of couples in societies just emerging from a traditional state, where most actions are specified by social custom and cultural arrangements of long standing – behaviour requiring sustained action by pairs of people on a matter of the utmost privacy and delicacy, plus the complications provided by one of man's strongest drives, sex.

The array of obstacles to be overcome before achieving success is impressive: illiteracy; ignorance of family planning methods, purposes and consequences; inertia and apathy, both general and specific; peasant resistance to change; costs to both the society and the individual; dispersal of population; lack of communication between husband and wife, both general and specific; the personalised character of the subject; the desire for children and/or for sons for political, economic and status reasons; occasional moral, religious or ideological objections; early marriages and high marriage rates; lack of alternatives available to women; occasional ineffectiveness of the simpler (hence more usable) contraceptive methods that discourages at the outset; lack of trained personnel for action programmes; little differential fertility as a lever to start with, that is, few small-family models; lack of an adequate system of distribution of supplies; remote and problematic rewards for successful action; invisibility of social support, deriving from the privacy of the practice. These are some of the salient obstacles that have to be overcome. On the other

hand, we can take heart from some favouring factors: pressure from the sheer struggle for subsistence, which gives a motive; governmental support, or at least permissiveness; general social change in a direction favourable to fertility limitation; improved contraceptive methods; and, finally, a widely professed interest in family planning, especially among the high-parity couples (though as a leading demographer once reminded me, that finding is based on an abstract answer to an abstract question).

I have found the following statement by Frank Notestein a dramatic way of thinking about the problem of effecting family planning, as against effecting some other health innovations of benefit to the society. 'Consider what would happen,' he once suggested, 'if malaria were as welcome as children – if a majority of young couples felt that they had really not justified their existence until they had undergone four or five attacks of malaria, which, moreover, they thoroughly enjoyed; if their fathers, mothers, mothers-in-law, uncles and aunts were constantly urging them to become exposed to the disease as soon as possible: if each new onslaught were welcomed with approbation by the whole community; and if to avoid this attractive disease, each deviant couple had to spray its own home with DDT acquired somewhat furtively.'

Now for the ten points I wish to make.

II

1. Three factors

I've come to think that there are three factors, or clusters of factors, that are involved in the effective spread of family planning. Our job is to learn as much as we can about each of the cells in the matrix that these three factors, together, make up. They are (*a*) the nature of the society, from traditional to modern or modernising; (*b*) the nature of the contraceptive method, from hard to easy; (*c*) the nature of the communication approach.

(*a*) *Nature of the society.* In a rough way, we can order societies, communities within societies and individuals within communities from the most traditional to the most modern, or modernising. By traditional we refer to much of the so-called underdeveloped countries, the countries that give an inferior role to women, that have a low level of literacy, that have little industrial development, and mainly an agricultural subsistence economy. By modern or modernising, we mean an approximation to the present conditions in Western Europe or Japan – high literacy, high degree of industrialisation and so on. This can get terribly complicated, but I think for our purposes it does not need to be. Generally speaking, some countries are more modern than others, some communities are more modern than others (for example, the cities within the underdeveloped countries are usually more modernised than the rural areas) and some individuals are more modernised, by this definition, than others (for example, the better educated). And in general we know pretty well that family planning, like any innovation, goes better at the modern end of the continuum – where there is more flexibility, more fluidity, less hold of custom, more literacy, more exposure to new ideas, more ideas around, more innovating leaders as models and so on. *We need not do more studies to find out that family planning will do better at this end than at the other end, but we do need to know how far down this continuum we can have some effect and at what cost.*

(*b*) *Contraceptive methods.* We might also order contraceptive methods on a continuum, but, for present purposes, it is sufficient to dichotomise them. Let us simply classify them into 'hard' and 'easy' on the basis of whether or not sustained motivation or repetitive action is required. By 'hard' I mean those contraceptives that require sustained motivation and repetitive action, usually before the very act of intercourse itself. It is relatively hard to teach or to use the rhythm method, diaphragm, condom, foam tablets, etc. By 'easy' I mean the newer methods of contraception that, in one way or another, by-pass this problem of sustained motivation and repetitive action. To a certain extent the oral pill does this. It does require some amount of sustained motivation; but it doesn't require action prior to the sex act itself and it's a relatively easy act that one member can do privately. But it does require supplies – thus you have the whole logistics problem – and it does require regularised attention and 'doing something'.

The newly developed intra-uterine devices are more nearly ideal from the communication and motivation standpoint, although all of the bio-medical evidence is not yet assembled. The devices are inserted into the uterus by a simple action that takes only a few minutes. They are, so to speak, a one-shot contraceptive method. If accepted and retained, this device works indefinitely, so far as we know. So far it has remained in some uteri for a year or two

years without any problems. This device by-passes the tremendous problem of sustained motivation which I think is close to being a key to the effectiveness of any programme. If the couple does not want to have children for the next year or so or longer, it can accept this device and from then on nothing needs to be done. It is reversible; if the device is removed, the women can become pregnant immediately. It has many of the characteristics of what people call the ideal contraceptive. These comments are not made as a sales talk for intra-uterine devices. The only point I want to make is that with regard to the implementation of family planning programmes we are in a completely different situation if we have at least some methods to offer that do not require sustained, repetitive actions, especially just before or as an interruption to the sex act.

(*c*) *The communication approach.* By this I mean the way in which informational and educational programmes are put into the field. Before expanding on this point, let me first note that *at any given point in time the first two of the three factors are pretty well fixed:* the society and the technology are given, and there is not much you can do about them but accommodate your programme. *The communication approach, however, is not given.* That is up to us. There are a large number of ways by which we can think of taking this programme to the people. We can choose to go to the people directly or to work through the community leaders. We can choose to make family planning part of a total programme of socio-economic improvement or we can choose to try to organise communities for family planning specifically. We can choose to make extensive use of mass media or we can try to inform people through small group discussions, or both.

One of the most useful choices we can make is in deciding where to concentrate our efforts and what methods we will advocate. Perhaps this point can be made more clearly if we construct a cross-tabulation of the first two factors and then consider the alternative communication approaches that could be made. The simple fourfold accompanying tabulation may be used to represent the first two factors.

	Type of society	
Type of methods	*'Traditional' societies*	*'Modern' societies*
'Hard' methods	Impossible	Possible
'Easy' methods	Possible	Possible

Some people think it isn't worth going into family planning in the most 'traditional' communities because you must introduce many other changes before you can effect family planning. They assert that you need to have some small amount of literacy, to free the women from their low status, to get industrialisation going, to start community development, to succeed in getting the adoption of improved agricultural practices, before you can get family planning to take root. In other words, they believe that most or all of the basic elements of modernisation must already be present before a communication programme can have any significant effect. This is an important research question. Can we do anything at all about family planning at the traditional end of the continuum? If research shows that very little can be accomplished here, then it may well be that we should not try – because we don't have enough resources anyway. If research supports this conclusion, we should concentrate our efforts towards the modern end of the continuum – and every society does have communities and groups and individuals at that end of the continuum.

The type of society may influence the methods we choose to communicate. It may be, for example, that in the traditional societies the 'hard' contraceptive methods are very difficult to get adopted – perhaps impossible. The type of society may also influence the type of communication approach we should use. For a full scientific understanding we need to ask and answer, by research, all of the combinations of question provided by the model. But we can't do this. It is impossible to undertake all the studies needed to accomplish this, but at least it is useful to keep this matrix in mind, and to see where various studies fit and what we are learning about the traditional types of society as against the more modernised types. We need to ask ourselves a question like, 'Does the same amount of effort or cost give us twice as much success in one type of community as it does in another, given the same methods and the same approach?' If so, that's an economic fact that the administrators of family planning programmes can take into account in deciding how they should allocate their total effort.

In this connection *a question of extreme importance is whether family planning can move ahead of the general modernisation of society.* My

personal opinion (and I think it is pretty clear that I'm in a minority on this) is that it can. My feeling is that in the general movement of modernisation the various components don't move ahead in direct synchronisation according to a single unvarying pattern. A 10 per cent improvement in education doesn't become a 10 per cent improvement in agricultural innovation (or any other fixed ratio). The components are uneven and do not progress in any set pattern. So *something has to be in front, and why not family planning?* I think that given present motivation among the high-parity couples (where professed desire for fertility limitation is generally strong), given an easy and effective contraceptive and given a reasonable system of information and services, then family planning can indeed be effective even in the 'traditional' societies – well in the lead of the modernising movement. But, as we said, that is a research question and one of the things we ought solidly to know in order to guide the administrative organisation of these programmes.

Basic research on this question is needed to attain the goal of *feasibility*, mentioned in the introductory remarks. We ought to tie our studies to realistic administrative decisions that the policy-makers will have to make. In doing this we should be, so to speak, researching the future, because the future is going to be different from the past and the present. The future is going to have more easy methods, and whether that future is three years from now or five or seven or ten, it certainly is coming. In any case, studies always take a long time to do; if we start a study today, we won't get the results (especially on this matter where you have to wait for the birth rate to respond) for two years or more.

With respect to the various methods of communication approach, it is my suspicion that our research is going to show that *any* reasonable approach is about as effective as any other. Given some information about methods and given the supplies, couples who are motivated to fertility control, usually by high parity, will undertake the practice. Couples who are not so motivated probably cannot be influenced to adopt family planning by direct communication under any circumstances, in the short run. If this is so, then the cheapest approach to the high-parity, favourably disposed and widely scattered minority is the best one, at least at this time. These remarks should make it clear why I regard the factor of method as being so important. There's an inverse relationship between the amount of motivation required for doing family planning and the ease of the method available. If there is an easy method available – one that requires no sustained action, little learning or skill, and that is very effective, cheap, non-troublesome, non-bothersome, not unaesthetic and so on – then you don't need very much motivation. But if you have something that requires persistence, skill, learning, that needs supplies, that costs money, that's unpleasant, that's objectionable to one or the other partner, then you need a great deal of motivation to keep that up. Usually there isn't that much motivation available, immediately. If you get enough motivation, of course, you can use almost anything and reduce the birth rate successfully. Historically, Western Europe did. Western Europe brought down the birth rate before any of these devices were available. But it took a very high level of motivation, and it also took a long time. With easier methods, we should be able to get faster results with much less motivation, and in traditional societies.

2. Short run

To make sense in talking about the implementation of family planning, we have to put a time scale on it. Most family planning programmes and research have a short-run (five to ten years) time scale in mind. What can we do in five to ten years? We should be doing research on both long-run and short-run time scales. But we need to emphasise achieving success quickly, because if we don't reduce fertility in the next ten years, the problem is going to be much worse in the next twenty years and so on, becoming progressively worse. We need to have an effect as soon as possible.

With regard to the long time period, here again I think I hold a somewhat deviant position. Many people feel that this is a terribly complex, complicated and difficult problem, and it is. However, I want to warn against accepting a cliché that academicians who aspire to do something of practical use often state. This is the claim that you need *full* understanding based on a long-run programme of 'basic' research before you can successfully deal with a social problem of this sort. This is a useful cliché for some purposes, but let's be clear that it is a cliché. We regularly take action successfully on all sorts of social problems without having full knowledge of 'how it works'. So let's not feel that we have to

understand the world population problem in any complete sense before we can act on it. In the intra-uterine device, incidentally, we have a very good example of this point. It is almost the most effective contraceptive method there is. Yet as of today the bio-medical people do not know its exact mode of operation: scientifically, they don't know *how* it works. But we can use it to great effect, nevertheless.

Similarly, we don't have to have a full understanding of motivation and communication for family planning in order to be able to get a successful programme going. We don't have to understand all there is to know about the cultural, personal, social, situational, economic factors and so on. Fifty years from now maybe we'll understand what we were doing that was effective, but we can do a lot without having complete understanding. From a long-run point of view that demands complete and detailed knowledge it is a very complicated problem; you have to understand everything before anything can be done. This calls for tremendous amounts of basic research. On the short time span you don't need that much knowledge or so much research. In the short run, you recognise that there are some people who are more highly motivated for family planning within any particular society and some who are less motivated or not motivated at all.

As I read the literature, there are two factors that strongly correlate with motivation and that go a long way towards explaining its level. One is *social-economic* status – class, educational level, standard of living and so on. I think that almost every study will show that the better-educated people are more receptive to family planning than the lesser educated; the wealthy more than the poor; those in higher occupations more than those in lower occupations. The second factor is *high parity*. By high parity I mean three or four children, of whom at least one is a son. The studies all show that if you ask people a question such as, 'Are you in favour of birth control?' or 'Would you want to practise birth control?' the curve of acceptance goes up with parity. Couples with no children or only one child have very little interest, those with two children are beginning to be a little interested and by the time you get to those with three or four children (including a son or two) you have high motivation. Those people are ready to learn about family planning, and my view is that in the short run, the best thing we can do with our limited resources is to satisfy the existing motivation. To put it in marketing terms, the thing to do is to exploit the existing market instead of trying to expand the market. You can't exploit the existing market fully anyway, because there are not enough resources. So this is the most efficient thing to do in terms of allocation of scarce resources.

Exploiting the existing market is the strategic thing to do for another reason; it is the best way to spread the market, to make family planning more in demand. Let those who are strongly motivated 'sell' the idea of family planning to those who are only mildly motivated. Gradually the practice will infiltrate the entire population. To get a programme started, we must develop a body of social support via satisfied users. Bringing family planning promptly to the few who are motivated now is probably the best thing that can be done to spread awareness, interest and desire to try it. Now this means 'skimming the cream' in a motivational sense – that is, deliberately planning a communication programme that will influence only 20 to 35 per cent of the society within a reasonably short time.

3. Implementation, not persuasion

As I mentioned, I came to this post as an alleged specialist in persuasion, but I quickly persuaded myself that persuasion wasn't the problem. In a way this is an amusing situation. The experts in communication say that the important thing is to get easy methods; this redefines the problem as a bio-medical one. Yet the doctors go out into the field, and they come back saying it is completely a motivational problem. They report, 'We gave people the method and they wouldn't use it, so why don't you fellows in motivation get their motivation higher?' Everybody is pointing the finger of responsibility at somebody else.

The major need, in my opinion, is to furnish information to the people who are uninformed but interested. A great deal more emphasis needs to be placed on the *administrative* and *organisational* aspects of family planning services. If there is one deficiency in this field that is more serious than any other, I think it's neither lack of an ideal method nor inadequacy of our persuasion techniques. *It is a general lack of adequate implemental machinery.* We don't have family planning administrators who know how to furnish services efficiently to great masses of population. We don't have a cadre of people who can organise a governmental programme, for example,

and apply it to a large population and run it successfully. I believe that the desire for fertility control (motivation) is high enough to give us a good, strong beginning in most of the nations of the world – if we could satisfy it properly. Information is low; practice is lower; we don't have the network of administration and organisation to take this programme into the field. We need to build it.

4. Bureaucratic timidity

There is one important exception to the comments just made concerning persuasion. Persuasion is needed at one point in the society: governmental officers are badly in need of it, especially very high government officers. The people themselves, I think, need less persuasion than the public servants (especially at the elite level) who have ultimate responsibility for family planning programmes. There are programmes where the response of the general citizenry seems very gratifying but where the expansion of the programmes to larger populations is inhibited by lack of support – real support as against verbal expression – at governmental levels.

There may be two good methods of persuading government leaders. Getting the true demographic situation and its implications brought to the attention of responsible governmental officers is one of the best things we can do at this point. Another avenue is to conduct surveys of attitudes towards family planning which measure desire for information and service, and bring them to official attention. Public presentation of the facts showing both how essential it is to the nation's future and how strongly the people sense the need for family planning and would welcome it in their own lives can encourage them to take a stand. If there is to be any success, it will require that responsible government officers get behind the family planning programmes in sufficient force to carry them out fully. If the people running them are always concerned about how their superiors are going to react or are handicapped by recognising ambivalence in support or are frustrated by the usual difficulties, then the already hard problem becomes harder still.

To make this specific, let's take an example. A major resource for the spread of family planning in most countries – and a neglected one – is the men under draft in the armed services. Most countries have large numbers of young men aged 17 to 25 who must spend two years or so in the armed forces. After military service they go back to their villages or their homes in the towns and cities. Now wouldn't it be the most natural thing, if you were seriously interested in effecting family planning, to have a major educational and informational programme for those young men before they go back? They are at the beginning of their own marital careers; they are going back into the whole stream of the society; they're going back in sufficient numbers to reinforce one another when they are back in the villages; they are, in a sense, a captive audience. While they are in service there are other kinds of instructional and informational programmes for them, for example, literacy programmes. Why shouldn't they all be made literate about family planning in order that they could begin to spread the idea and support it back home? In order to do that you obviously have to get strong support at the highest levels of the armed forces. There are no extraordinary problems that need to be solved to carry out this programme; we know how to do it. What is needed is permission and genuine support, i.e. the opportunity to put efficient programmes like this into effect.

Very prominent in the elite that needs persuasion is the medical profession. The medical community is an indispensable partner to this whole enterprise. The plain fact of life is that in almost all countries family planning is seen as a health problem. The medical community will make it a health problem, and you can't get anywhere with family planning without the help and co-operation of medical men. They should be involved in these programmes from the outset, and thoroughly.

5. Mass media as legitimators

What is the role of the mass media in family planning? By mass media I mean newspapers, radio, pamphlets and so on. (Television doesn't exist in most of the developing countries.) Here let me quote a paragraph or two from something I wrote earlier:

> With regard to objectives, the job of the mass media is twofold but the second task is unappreciated or insufficiently appreciated. The one task of the media, the one that is strongly recognised, is to generate interest, provide information, change attitudes, stimulate action; that is, to do all those things that a programme of information and education is supposed to do. All

of us believe, I suppose, that people should be allowed to make their own decisions in matters affecting their own welfare and involving how they will spend their time and money, but they should make them on the basis of good and full knowledge. The media are supposed to provide that knowledge. Although the term is suspect in some academic circles, we can call this the 'rational' use of the media. The other function the media can play with respect to a new idea or a new practice is to help make it look *legitimate*, to help make it familiar and acceptable. This is a most important and worthy task, and if anything perhaps more effective than the rational action function. The job is to develop the perception of social support for an innovation within the community. How can that be done? By spreading the perception that 'other people like me are doing it and deriving benefit from it'. This is something the mass media can do, and it is critical for the acceptance of almost any conceivable programme. It is neglected, I fear, relative to the rational use of the media.

If one generalisation emerges as much as any from the communication studies done on change of opinion over the past years, it is that people are strongly influenced in attitudes and behaviour by other people, by their direct personal contacts and by their perception of what they consider the others consider the right thing to do. In short the job of the media is, in this sense, both rational and social. What this means in practical terms is that, when devising informational and educational material, we remember that in addition to the content itself, one should include material to show that this is socially acceptable and personally beneficial behaviour – that doing what is suggested is taken for granted as a natural part of a good and wise life by people as close to the subjects of the informational programme as one can honestly get. The very massiveness of the mass media can promote this image of social acceptability. The job of the media, then, is not simply to spread information but to spread a sense of social rightness, even righteousness, about the practice.

This needs to be underlined because it is so important for the developing nations. It's difficult for some of us to appreciate how novel family planning can seem to a population that hasn't been particularly concerned with it. And with anything that is novel, people aren't quick to pick up the innovation unless they know that it is approved or taken for granted by the people around them or the people whom they esteem.

In this regard, family planning has a handicap which most of the innovations do not have. If you can develop a better way to raise rice and get twice the yield as traditional methods, the benefit is tangible and quickly visible. But family planning is invisible. It is private in the sense of being personal, and it's also private in that it isn't talked about very much; it isn't a 'normal' topic of conversation. As a consequence, it needs to have social support built up much more than most innovations. This is one job the mass media can do. The mass media can combat what the sociologists call 'pluralistic ignorance', a situation where people favour an idea but think everyone else is against it, and so are afraid to speak or act in an approving way. The media might be able to turn that around, and if so they will have done an important job.

6. Communication not refined

In my opinion, one doesn't need a highly refined, a highly nuanced presentation of the case for family planning. In helping to design a communication programme for Taiwan, I reviewed the literature to see what principles and what theories should be incorporated in these materials. After a while I came to feel that this search was a little silly. The messages we have are few in number; they are not difficult ideas, and they seem to be pretty much the same the world over. The people need to be told 'what' (the elementary facts about the physiology of reproduction), 'how' (contraceptive methods), and 'why' (the rationale *for themselves*). How many reasons can there be for practising family planning? The number is not infinite. There are only about five or so important ones. Whether you have a blue cover on the leaflet or a red cover doesn't make any significant difference.

We can be misled in thinking about communication if we generalise from our experience in this country. We live in a very atypical situation. We're bombarded by communications all the time. Some people want us to buy Gleem toothpaste, some Crest and some Colgate: so whether the package is red or green might really make a difference to the people who like red better than green. We live in a highly competitive communication situation and

every little refinement might make a per cent difference in the market, which could make a great deal of difference on the profit side.

But on this topic and in the areas we're mostly concerned about, there isn't that kind of competition from the communication media. Quite the contrary; there is hardly any mass communication at all, so that anything you say is attractive just by virtue of the fact that it is novel. It has a kind of entertainment value, a stimulus value, just from the fact that it is there, that it exists. Once we've told the basic story of 'what', 'how' and 'why', with some personalising humour and a little entertainment, once we've done that and repeated it ingeniously, then we've just about done what we can so far as the sheer communication context is concerned – plus, of course, doing what we can to build the message into the social framework.

7. Success is not 'success'

Here I want to make a plea for realism. What is success in this field? Let me emphasise that you are successful if you can move a relatively few people. That's what success is. Success in family planning can never consist in getting an entire community to accept this practice, certainly not in our short run of five or ten years.

Let's start with the following enlightening, and sobering, set of data. Let us imagine that we start to promote family planning in a given community, and the target population is all married women from ages 15 to 45. This is equal to 100 per cent. When you go into the community, this is what you find: (*a*) First, x per cent are already using something, so that's fine; you don't have to worry about them, except that what they're using might be less effective than what you think you have to offer. (*b*) Then, y per cent are sterile, or think they are, so they're not available for your programme. (*c*) z per cent are pregnant or lactating. If they are pregnant, they don't need family planning, at least not now; and if they are lactating, they think they don't. Now if you add up these three figures – the total varies, of course, from population to population – they can amount to as much as half the population, and seldom less than a third, unavailable before you begin. (*d*) Then there is the group who positively want more children now, or who want a son (typically people of low parity). And how many are they? Perhaps another quarter. So as many as three-fourths of the married women may be out-of-bounds with respect to the programme before we begin. One-fourth of the initial 100 per cent are what you might call currently eligible at any given interval of six months or a year. So, if we can be successful in getting our message adopted by 50 per cent of the currently eligible (a very high return), that means we would get only 12·5 per cent of the original target. In this example success, very great success, would be 12–13 per cent. We'd be satisfied with about one-half of that or 6 per cent. Now the point of what I am saying is that 6 per cent is success, it isn't failure; 6 per cent of the 100 per cent is success, not failure, in building family planning practice.

In a way there's a certain arrogance in our implicit notions, sometimes explicit, that people should give us pieces of their attention, their minds, their time (sometimes their money) precisely when and because we are sure we have something good for them. I say this with a sincere feeling that what we have is indeed good for them, but I still find it humbling to keep in mind Thoreau's observation that 'When I see someone coming towards me with the sole aim of doing me good, I run for my life'. People have their own lives to live, and they are being bombarded from all sides by other people who are also convinced, no less firmly than we, that they have something essential to offer. Even if everyone were willing to give serious attention and serious efforts to all of these opportunities (a dreadful vision in itself), there may not be enough attention to go around. And that is part of the reason why we have to settle with a small fraction and call it success.

As realists with a difficult task, let us maintain professional expectations lest we start blaming the people for being blind to our attractive wares. They are the ultimate judges of the attractiveness. It is not unrealistic to assume that for the first few years of new family planning programmes, success will be far closer to 1 per cent than 100 per cent of the couples. But as I suggested earlier, this proportion of initial success, if not infinitesimal, has a built-in chain reaction of its own. Happily, from our standpoint, this chain reaction not only is quite effective but also free. Less than complete success, then, may be a cause for regret but certainly not for despair or disillusionment.

8. Cost

I'm coming to believe that we are moving from one slogan to another. The former slogan was, 'It can't be done.' There were a lot of early failures in trying to effect family planning, and people in the field said, 'If we could only get a success somewhere to point to!' The Singur project was one success; we think we now have others in Singapore, Ceylon, Taiwan. So I think we are moving from that slogan to the current slogan, 'We can do it but it costs too much.' I think that we do have successes to point to, but they are too expensive. We have to find ways to spread birth control more cheaply. The newer methods will help, and so will the mass media, and so will word-of-mouth, but beyond that I have little more to suggest.

9. Place of research

Where does research come in? Briefly, my feeling is, again thinking of this short-run period, that basic research, as it would be defined by many a sociologist or psychologist or anthropologist, is not warranted, because it will not yield enough results to be helpful in the short run. Doing academic studies on family structure, or community power structure, or innovative personalities, or communication networks, I think is not going to pay off with results in time to be useful. I think that the role of research is the applied role. I don't use that term invidiously. I'm proud to be an applied social researcher. The research needed is primarily in the evaluational role, appraising the results: 'How much effect are we actually having in what we are doing, at what effort, how and why?'

Let me give an example that I think is a fair one. The noted psychologist Carl Hovland, at Yale, became famous for doing experimental studies on communication effects. He studied communication effects by laboratory experiments carefully controlled, very nicely done – and without doubt he was a major figure in this field. One of his studies, done during the war, was a so-called one-sided, two-sided argument study. It was devoted to the question, 'In trying to persuade people, should you present only one side of the argument or should you present both sides – the arguments against your position as well as those favouring it?' Hovland did a whole series of experiments on one-sided and two-sided persuasion. The results were quite clear: (*a*) with people who initially disbelieve what you are trying to tell them it's better to give them the two-sided argument so that they know that their arguments were taken into account and answered; (*b*) educated people want two-sided argumentation more than uneducated people. Now how can we apply the Hovland experiments to family planning? In my opinion we cannot make much use of them. In the first place, on ethical grounds you should present both sides. You don't have the right to put together materials like this, certainly not for another society, and just present a one-sided case as though the other side didn't exist. Secondly, just on the basis of common sense and any experience at all in the field, you recognise that while these people may be illiterate and may be uninformed, they are by no means unintelligent. If you don't speak to their concerns, you might just as well not speak at all. So you have to use the two-sided approach on both ethical grounds and common-sense grounds.

What I think we need from research is the careful evaluation of results. From the standpoint of the scientist, you do research only where you don't know the answer, and you'll accept the answer, whatever it is. That's what research ought to be. Once you have stated the question and you have gone through the inquiry with proper methods, you are committed as a scientist to accepting the results, however they turn out, whether you like them or not. In this field, I think there is one result that research can give that will never be acceptable, certainly not to the administrator and not even to the scientist, and that is a negative answer. No amount of research showing that family planning cannot be done is going to stop the effort from going forward. There is just too much importance attached to it, too much investment in doing it – economic, social, psychological. But research can never give a fully negative answer anyway. What research can do is to eliminate needlessly futile efforts. If we do research in connection with the three factors of my first point, we can narrow the range of the worth while and delineate the realm of failure and the impracticable.

10. Direct approach

How then should we go about implementing family planing in the short run? My answer is, 'The way to do it is to do it.' We shouldn't back off and say that we have to have endless basic

research before we know how the family structure ties in to the community leadership. There are so many limitations on what you can do in the field any way. You don't have a *tabula rasa*; you have only a few limited choices you can make. The best thing to do is to move into a situation, set up as reasonable an experiment as you can to test the elements of choice you do have. Set up the implementational machinery, train the family planning administrators and field workers and start operations. Adjust your efforts as the appraisals come in, remembering that if you get 6 per cent you've had a success, not a failure, and continue. Keep a sharp, commonsense eye on the effort as you proceed, and keep yourself free to shift gears as the terrain indicates. Get the implementational machinery and start. Experiment as much as you can; see what works and what doesn't; hope that the bio-medical people will produce an ideal contraceptive before you are very far down the road. *And evaluate the results firmly as you go.*

III

My concluding impressions, then, are (1) that in the typical traditional community of the underdeveloped countries there is a wide range of motivation for and saliency of family planning; (2) that a substantial group at the lower end of the range cannot be motivated, at least not in a period of five years or so, but groups at the other end, especially the high parities, are ready now, a group representing from one-fourth to one-third of the community; (3) that given methods of good effectiveness and not too much bother, such people will use family planning with satisfactory consistency, enough to make some difference in the birth rate; (4) that the best way to motivate new users is to satisfy the motivation that already exists, that is, the best way to spread the market is to exploit the available market; (5) that the people are ill-informed about family planning but by no means unintelligent about this or other matters affecting their lives closely; and (6) that given this state of affairs any reasonable programme of communication plus supplies will do about as well as any other; certainly it will not require a high degree of nuance and refinement in the campaign to distinguish between success and failure.

NOTES ON EDITORS

James Barber, M.A., Ph.D.
Reader in Political Science

James Barber was with the Uganda Civil Service for seven years and subsequently taught at the University College of Rhodesia and at Exeter. He is particularly interested in African politics and foreign policy, and his publications include *Rhodesia: Road to Rebellion* and *Imperial Frontier: A Study of Relations between the British Administration and the Pastoral Tribes of North-Eastern Uganda.* He is currently working on a study of Anglo-South African relations since 1945.

John Blunden, B.A., Ph.D.
Senior Lecturer in Geography

John Blunden worked for six years as a B.B.C. Producer before taking up an appointment as Lecturer in Geography at the University of Sussex and Fellow at the Centre for Educational Technology. His chief research interests are in agricultural geography and land-use planning. He is at present completing a book on Britain's extractive industries in relation to land-use planning problems.

F. S. Brooman, M.A.
Professor of Economics

F. S. Brooman was Senior Lecturer in Economics at the University of Bristol before joining the Open University. His particular interests are in the fields of money and banking and the economic problems of Latin America; he is the author of *Macroeconomics.*

Hedy Brown, B.Sc. (Econ), M.Sc.
Lecturer in Social Psychology

Hedy Brown was previously a Senior Lecturer in Social Psychology at the North-Western Polytechnic. Her particular interests are social change, the social psychology of institutions (particularly hospitals) and educational problems.

Michael Drake, M.A., Ph.D.
Dean and Director of Studies in Social Sciences

Michael Drake has taught at Trinity College, Dublin, Queen's University, Belfast, and most recently at the University of Kent. Professor Drake's research interests are in historical demography and historical sociology. He is the author of *Population and Society in Norway, 1735–1865.*

Ruth Finnegan, M.A., B.Litt., D.Phil.
Lecturer in Sociology

Ruth Finnegan's previous experience lies mainly in Africa, where she has held teaching posts in the University College of Rhodesia and Nyasaland and the University of Ibadan and carried out research in Sierra Leone and Nigeria. Her current research interests are in the sociology of oral literature and oral communication (in both industrial and non-industrial societies). Her publications include *Oral Literature in Africa* and two books on the Limba people of northern Sierra Leone.

Robert Green, B.Sc., Ph.D.
Professor of Psychology

Robert Green was a Lecturer in Psychology at University College, London, until 1969. He is particularly interested in selection procedures, personality measurement, memory and perception and the measurement of political imagery; he also spent a year on secondment to the Government of Trinidad and Tobago as an expert in educational testing.

Alan Hancock, M.A.
Senior Producer, B.B.C.

Alan Hancock is the Senior B.B.C. Producer working with the Social Sciences Course Team. He has worked for a number of years in the educational television departments of the B.B.C. and recently spent two years in the Far East working as an adviser on educational broadcasting projects. His publications include *Mass Communication, Television and Social Work* and *Planning for Educational Television.*

A. T. A. Learmonth, M.A., Ph.D.
Professor of Geography

Andrew Learmonth was previously Professor of Geography at the Australian National University in Canberra and his particular interests are medical geography and regional planning in India. His publications include joint authorship of *India and Pakistan*, *The Eastern Lands*, *Encyclopaedia of Australia* and *Regional Landscapes of Australia*.

D. J. Murray, M.A., D.Phil.
Professor of Government

David Murray has recently been teaching and conducting research in East, Southern and West Africa. His principal interests are in the fields of comparative administration and African government and politics. His publications include *The West Indies and the Development of Colonial Government*, *Work of Administration in Nigeria*, and *The Governmental System in Southern Rhodesia*. He was also editor of *Studies in Nigerian Administration*.

Jeremy Tunstall, B.A.
Senior Lecturer in Sociology

Jeremy Tunstall has held research posts at Hull, the London School of Economics and most recently at the University of Essex. His main publications are *The Fishermen, Old and Alone* and *The Westminster Lobby Correspondents*. He was also editor in 1970 of *Media Sociology*.

INDEX OF AUTHORS

Chapter	Page	Name
19	293	Alexander, J. W.
25	459	Almond, G. A.
22	362	Ambrose, P.
	371	
	374	
	393	
31	578	Asch, S. E.
6	121	Beach, F. A.
36	684	Berelson, B.
21	351	Berry, B. J. L.
22	366	
	375	
7	143	Bettelheim, B.
8	160	
23	409	Blij, H. de
26	472	Bohannan, P.
22	385	Bracey, H. E.
4	50	Brookfield, H. C.
	65	
22	376	Brush, J. E.
	385	
21	349	Burgess, E. W.
	350	
22	393	Bunge, W.
29	599	Cohen, P.
20	329	Colberg, M. R.
30	574	Coser, C.
24	424	Dahl, R. A.
2	23	d'Entrèves, A.
	27	
	30	
	35	
	35	
6	109	Dollard, J.
27	491	Easton, D.
1	15	Erb, E. D.
22	368	Everson, J. A.
22	368	FitzGerald, B. P. F.
6	121	Ford, C. S.
28	535	Frankenberg, R.
35	655	Freedman, R.
10	204	Galbraith, J. K.
21	341	Garner, B. J.
22	375	Garrison, W. L.
9	199	Gavron, H.
33	620	Geddes, A.
20	329	Greenhut, L. M.
19	289	Grötewald, A.
20	318	Haggett, P.
21	358	
20	335	Hamill, L.
20	317	Hamilton, F. E. I.
	330	
21	357	Harris, C. D.
21	351	Hartmann, G. W.
15	256	Hicks, J. R.
19	297	Hidore, J. J.
1	15	Hooker, D.
23	413	House, J. W.
21	341	Hoyt, H.
Figures passim		
8	160	Janowitz, M.
14	246	Johnson, H.
8	168	Katz, D.
36	674	Katz, E.
8	154	Kelman, H. C.
22	384	Kohn, C. F.
33	637	Learmonth, A. T. A.
34	648	Lewis, W. A.
27	513	Lipsitz, L.
26	477	MacCallum, S.
8	168	McClintock, C.
23	394	Mackinder, H. J.
20	338	Macmillan, T. E.
34	643	Malthus, T. R.
26	481	Mann, L.
20	329	March, J. G.
10	203	Marshall, A.
5	93	Marshall, L.
13	229	Matsushima, S.
22	384	Mayer, H. M.
30	567	Merton, R.
28	525	Michels, R.
31	594	Milgram, S.
6	109	Miller, N.
34	645	Myint, H.
25	440	Neumann, S.
31	585	Newcomb, T. M.
16	264	Nicholson, R. J.
12	223	Nove, A.
7	138	Ogburn, W. F.
18	276	Pear, T. H.
19	313	Peet, J. R.
12	215	Phelps Brown, E. H.
22	366	Pred, A.
3	42	Radford, R. A.
14	250	Robinson, J.
13	235	Roy, D.
32	611	Ryder, N. B.
8	168	Sarnoff, I.

Chapter	*Page*	*Name*
24	428	Sartori, G.
11	212	Schumpeter, J. A.
21	340	Seyfried, W. R.
21	341	Simmons, J. W.
20	329	Simon, M. A.
3	41	Smith, A.
	42	
21	353	Smith, P. J.
	354	
20	319	Smith, W.
17	271	Srinivas, M. N.
9	183	Stenning, D. J.
20	332	Tarrant, J. R.
27	501	Tawney, R. H.
21	341	Tennant, R. J.
11	208	Turvey, R.
21	357	Ullman, E. L.
22	372	
	384	
	392	
9	175	Vogel, E. F.
19	289	von Thünen, J. H.
27	507	Weiner, M.
35	669	
5	80	Whyte, W. F.
12	215	Wiseman, J.
19	302	Wolpert, J.
21	342	Yeates, M. H.
	345	
	347	
	348	
	353	
1	3	Young, J. Z.

INDEX OF EXTRACT TITLES

Chapter	Page	
27	491	An Approach to the Analysis of Political Systems
22	362	Analytical Human Geography
8	168	Attitude Change: Procedures and Motivating Patterns
31	585	Attitude Development as a Function of Reference Groups: The Bennington Study
9	199	(The) British Family Today
21	353	Calgary: A Study in Urban Pattern
17	271	(The) Caste System in India
21	351	(The) Central Business District: A Study in Urban Geography
22	366	Central Place Studies
21	340	(The) Centrality of Urban Land Values
4	50	(The) Chimbu: A Highland People in New Guinea
21	349	(The) City
21	341	Commercial Structure and Commercial Blight
25	459	(A) Comparative Study of Interest Groups and the Political Process
24	428	Constitutionalism: A Preliminary Discussion
6	109	Critical Training Situations
19	302	(The) Decision Process in Spatial Context
10	204	(The) Dependence Effect
6	121	Development in the Individual
26	477	Dispute Settlement in an American Shopping Centre: A Preliminary View
16	264	(The) Distribution of Personal Income
3	41	(The) Division of Labour
19	293	Economic Geography
3	42	(The) Economic Organisation of a P.O.W. Camp
31	578	Effects of Group Pressure upon the Modification and Distortion of Judgements
8	160	Ethnic Tolerance: A Function of Social and Personal Control
20	329	Factors in the Location of Florida Industries
7	143	Feral Children and Autistic Children
1	3	From Molecule to Man
14	250	Full Employment
22	375	(The) Functional Bases of the Central Place Hierarchy
30	574	(The) Functions of Conflict
23	394	(The) Geographical Pivot of History
23	409	(The) Heartland Debate
11	208	How Rational Economic Decisions are Made
33	637	India's Population Geography in an Era of Change
20	332	Industrial Development in Ireland
3	42	(The) 'Invisible Hand'
19	289	(The) Isolated State
9	175	(The) Japanese Family
13	229	Labour–Management Relations in Japan
2	23	Law and Order
23	413	(A) Local Perspective on Boundaries and the Frontier Zone: Two Examples from the European Economic Community
20	318	Location Analysis in Human Geography
20	330	Location Factors in the Yugoslav Iron and Steel Industry
20	319	(The) Location of Industry
13	235	Making-out: A Counter-system of Workers' Control of Work Situation and Relationships
30	567	Manifest and Latent Functions

Chapter	*Page*		
20	317		Models in Geography
14	246		Monetary Theory and Keynesian Economics
15	256		Monetary Theory and History: An Attempt at Perspective
4	65	(The)	Money that Grows on Trees
2	27		Nature and Convention
21	357	(The)	Nature of Cities
2	30		Negative Liberty
35	655		Norms and Family Size in Underdeveloped Areas
36	684		On Family Planning Communication
34	643		On Population and Subsistence
21	341		One Hundred Years of Land Values
Figures passim			
20	329		Organisations
24	424		Political Systems: Differences
28	535	(The)	Politics of Recreation
34	645		Population Pressure in Underdeveloped Countries
32	616		Population Trends in Developing Countries
2	35		Positive Liberty
12	215	(The)	Pricing System and Resource Allocation
29	559	(The)	Problem of Order
11	212	(The)	Process of Creative Destruction
18	276	(The)	Psychological Aspects of English Social Stratification
1	15	(The)	Psychology of the Emerging Self
26	481		Queue Culture: The Waiting Line as a Social System
22	384		Readings in Urban Geography
19	297	(The)	Relationship between Cash-Grain Farming and Landforms
27	501		Religion and the Rise of Capitalism
32	611	(The)	Reproductive Renaissance North of the Rio Grande
22	385		Rural Service Centres in Southwestern Wisconsin and Southern England
22	368		Settlement Patterns
5	93		Sharing, Talking and Giving: Relief of Social Tensions among !Kung Bushmen
26	472		Social and Political Organisation of the Tiv
33	620	(The)	Social and Psychological Significance of Variability in Population Change
31	594		Some Conditions of Obedience and Disobedience to Authority
21	342		Some Factors Affecting the Spatial Distribution of Chicago Land Values, 1910–1960
12	223	(The)	Soviet Economy: Micro-economic Problems
19	313	(The)	Spatial Expansion of Commercial Agriculture
5	80		Street Corner Society
27	507		Struggle Against Power: Notes on Indian Political Behaviour
28	525	(The)	Struggle Among the Leaders Themselves
21	351		Studies of Highway Development and Geographical Change
20	335	(A)	Test of Weberian Location Theory
22	393		Theoretical Geography
22	372	(A)	Theory of Location for Cities
8	154		Three Processes of Social Influence
25	440		Towards a Comparative Study of Political Parties
36	674	(The)	Two-Step Flow of Communication: An Up-to-date Report on an Hypothesis
35	669		Urbanisation and Political Process
10	203		Utility and Demand
19	289		Von Thünen in Retrospect
34	648		What is the Right Size of Population in Relation to Resources?
20	338		Why Manufacturers Choose Plant Locations *v.* Determinants of Plant Location
9	183	(The)	Woɖaaɓe Today: Founding a Family and a Herd
7	138	(The)	Wolf Boy of Agra
27	513		Work Life and Political Attitudes: A Study of Manual Workers

GENERAL INDEX

Abipone, the, 129
Aboltin, report on Stalingrad construction trust, 225
Adaptation to demands of life, 8–9
Adjustment mechanisms, 21
Adler, Victor, 534
Adolescence and puberty, 123–4: adolescent sterility, 124
Advertising and salesmanship, 205–7
Aggression: and frustrated mobility aspirations, 120–1; devious, 119
Agra, wolf boy of, 138–43, 149–50
Air power, 407, 411
Alfieri, Vittorio, 25
Allingham, Margery, 280
Alonso, W., 340
Alorese, the: as semi-restrictive society, 131; puberty ceremonies, 126
Alpes-Maritimes boundary zone, 419–21
Alsace-Lorraine, Franco-German boundary in, 418
Amery, Leopold, 402, 406–8
Analytical Human Geography (Ambrose), 362–6, 371–2, 374, 384, 393
Andamanese, the, 131
Anderson, Nels, 200
Andhra: Formation of State, 511
Anger–anxiety conflicts, 118–20; devious aggression, 119; fear attached to anger cues, 118; mental limitations, 119; overinhibited persons, 119–20; patriarchal code, 118; sibling rivalry, 119; unlabelled anger conflict, 119
Anti-Negro attitudes, 161–3, 165-8, 170–3
Anti-Semitism, 160–3, 165–7
Antonov, O., 225
Apinaye, the, 128
Aquinas, St Thomas, 29
Arapaho, the, 129
Aristotle, 17, 35, 460: and *politeia*, 434, 436; classification of political systems, 424–6; 'naturalness' of the State, 27–9; on the State and justice, 27–8; view of democracy, 425
Arthasastra, the, 510
Ashanti, the, as restrictive society, 129
Asiatic influence on Europe, 395–400
Assam, 'colonisation' in, 622, 627–8, 633, 636, 640
Assembly-line workers compared with other occupational types, 514–22: economic preoccupations, 518–19; fatalistic attitudes, 517–18; jobs, 516; manipulative practices, 519–21; mental health, 521, 524; political radicalism, 518–19; sympathy and tolerance, 519; unions, 516–17
Attila, 397
Augustine, St, 24: definition of order, 26
Aurobindo *ashram*, 508
Australopithecines, 14
Authority and evaluation of knowledge, 16–17
Autistic children, 143, 145–53
Avars, the, 397
Aveyron, wild boy of, 143

Bacon, Sir Francis, 502
Badoglio, Marshal, 455
Bagehot, Walter, 441, 449
Bahaya, the, 656
Bakke, E. W., 277
Bank Charter Act (1844), 261
Banking control, 261–3: international aspect, 263
Baran, Professor, criticism of Keynes's theory, 253
Bebel, August, 527
Beer, Samuel, 444
Behavioural factors among farmers in Middle Sweden, 302–12: attitude to risk, 312; goal orientation, 306–7; knowledge situation, 307–8; normative and behavioural analysis, 303–6; productivity tables, 304–5; sample population, 302–3; uncertainty, 308–12; yield variability, 308–10
Bellers, John, 504
Bender Gestalt test, 173
Bengal: famine of 1943, 635; population changes, 622, 625–8, 635–6, 639–41
Bennington study of attitude development as function of reference groups, 585–94
Bereavement, response to, in India, 631–3
Berkeley, Bishop George, 506–7
Berlin, Sir Isaiah, 35, 37
Bernal, J. D., 4
Beveridge, Sir William (Lord Beveridge), on full employment, 252
Bhagavad Gita, 510
Bismarck, 533
Blackburn, J. M., 276, 277
Blake, J., 656, 658, 667
Bogardus Social Distance scale, 173
Bolingbroke, Lord, 434
Bormann, Martin, 452
Börne, Ludwig, 529–30
Bornu, State of, 183–4
Bossism, 568–70
Boundaries and the frontier zone, 413–23: case studies, 414–15; classification of boundaries, 414–15; context and perspective, 413–14; direct confrontation, 415; frontier zone, 416–17; life near boundary lines, 416–17; normal boundaries, 415–16; orthodox work on principles, 414
Brahmins, 271–4, 509, 510, 625, 627
Brainwashing, 155–6

Braithwaite, J. C., study of post-war industrial development of Plymouth, 330
Briand, Aristide, 532
British family today, the, 199–202: alleged instability, 200–1; basic functions, 199–200; working wives, 201
Broadcast Word, The (Lloyd James), 281
Brogan, Sir Denis, on social stratification in U.S.A., 276
Brown, Paula, 51
Bryce, James (Viscount Bryce), 441, 567
Buddhism, 274, 629, 631
Bulgarian invaders, 397
Bunke (branch family, Japan), 177
Bureaucracy, 449, 454: and interest groups, 469–70
Burke, Edmund, 434
Burma, population changes in, 629, 631, 645, 647
Burns, John, 532

Cain and Sheppard, study of small snails, 10
Calcul économique, le, 208
Calcutta, political protest in, 669–71: migrants, 670–1; refugees from East Pakistan, 670
Calgary, urban land use in, 352–61
California F scale, *see* F scale
Calvin, John, 504: Calvinist Churches in France, 503
Calwer, Richard, 535
Canada, fertility history of, 611, 615
Cash-grain farming and landforms in U.S. Midwest, 297–301; definition and distribution of cash-grain farms, 297–8; definition and distribution of flat land, 298–9; correlations, 299–301
Caste system in India, 271–5: concept of pollution, 272–3; diet and status, 273; fivefold hierarchy, 271; idea of hierarchy questioned, 274; *jātis* (sub-castes), 271, 272, 274; *kharma* and *dharma*, 272; movement between castes, 274; occupations and caste, 273; sanctions, 274; sanskritisation, 274, 275; 'twice-born' castes, 271
Casti, Giambattista, 535
Castration anxiety, 116–17
Catharsis, 171
Catholic Action, 457
Catholic parties and trade unions, 465–6
Cells, 6–7: basic simplicities of organisation, 6; structure of membranes, 6
Ceylon: and malaria control, 637, 645; population increase, 629, 648–8
Chagga, the, as restrictive society, 128, 129
Charismatic leadership, 449
Charlemagne, 529
Charles I, economic policy of, 502
Chewa, the: as permissive society, 133; initiation rites, 127
Cheyenne, the, 129
Chicago, land values in, 341–51, 353–4
Chimbu of New Guinea, the, 50–79: agriculture, 52, 56–60; airstrips, 64; changes since 1933, 52–3; climate, 54; 'discovery' of the Chimbu, 50–1; geographical situation, 50; habitat, 53–6; political groups and territories, 60–2; rainfall, 54–5; regional contrasts, 62–4; segmentary social structure, 51; soils, 55–6; suppression of warfare, 53–61; tribes, clans and subclans, 51–2
coffee-growing, 53, 59–60, 65–79: introduction and ecological adaptation, 68–9; expansion, 69; adjustment and conflict within farming system, 69–70; conflict of values, 70–1; victory of progressives, 71; centralisation of processing, 72–3; return to more dispersed processing, 73, 75; uncertainties, 75; evolution of central place, 75–8
China: and the Mongols, 399; Communists and brainwashing, 155; high fertility, 657; population density, 657
Chiricahua, the, as restrictive society, 128, 130
Chisholm, Michael, 297, 315
Chlorophyll, 5, 7
Christaller, Walter, central-place concept, 366–7, 369–76, 380, 393: fixed-*K* hierarchy principle, 369–71, 375; hexagonal lattices, 367–71; service centres and distance factor, 384–5
Christian Democratic movement, 457
Chromosomes, 8: mechanisms controlled by DNA, 9
Chuckchee, the, 129
Cicero, 28, 35
Clark, P. J., 382
Cleanliness training and conflicts, 111–14: advantages of verbal aids, 113–14; difficulties of learning, 112; escaping from parents' sight, 113; excessive conformity and guilt, 113; learning without verbal aids, 112–13; strong emotions aroused, 113; superego, 114
Coale and Tye, study of high fertility in China, 657
Cocos Island, population study, 655, 656
Coercion theory of social order, 560–2
Colbert, Jean-Baptiste, 502
Cole, Margaret and G. D. H., 276
Colic, 111
Columbia University, Bureau of Applied Social Research, 674, 684
Columbian epoch of history, 394, 401, 404, 407
Communication, two-step flow of, 674–84: Decatur study of decision-making, 674, 677–84; Elmira study of 1948 election campaign, 674, 677, 679–84; homogeneity of opinion in primary groups, 679–80; impact of personal influence, 679–80; new drug study among doctors, 674, 678–80, 682; opinion leaders and mass media, 681–2; personal and mass media influence, 679; Rovere study of interpersonal influence and communications behaviour, 674, 676–7, 680–4; voting study (1940), 675–6, 680
Communist Manifesto, 442
Communist parties, 443–4, 465–7: Comin-

tern and Cominform, 457; in India, 508, 513, 669, 670; Internationals, 456
Competition, types of, 213
Complexity of higher animals, 11–12
Compliance, 155
Comte, Auguste, 564, 565
Confucian ideology, 177
Congress Party (India), 508–13, 641, 670–72
Constant, Benjamin, 35, 37, 434
Constitutions: American tradition of 'limited rights', 434; as fundamental law in Western area, 429–30; basic elements 430; conventional approach, 432–3; façade, 434–5; *garantiste* feature, 430, 431, 434–6; non-fulfilment of constitutional provisions, 436; role in political system, 431; written and living, 435
Consumer demand and behaviour, 205–7
Cossacks, the, 401
Cost-benefit analysis, 208
Cost-effectiveness budgeting, 208
Costs, social significance of, 219–20
Council of Europe, 457
Creative destruction, 212–14: types of competition, 213
Critical training situations, 109–21
Croce, Benedetto, 34
Cro-Magnon man, 18
Crowd psychology, 450
Cuna, the, 128

Dangerous Thoughts (Hogben), 284
Danton, 526
Dart, Broom and Leakey, discovery of Australopithecines, 14
Darwin, Charles, 9: and survival of the fittest, 7; study of Galápagos Islands finches, 9
Davis, K., 656, 658, 667
Dean, W. H., 335
Decatur study of decision-making, 674, 677–84
Deccan, recurrent population crisis in, 622, 624–5, 627, 636, 640
Declaration of Right (1789), 434
Defence mechanisms, 21
'Deferential' democracies, 38
Demand: utility, 203; consumer demand, 205–7; law of, 204
Democracy: and dictatorship, 440–4, 446; and industrialisation, 522–4; and positive liberty, 36; 'deferential' democracies, 38; Rousseau's presentation, 37; threats arising from egalitarian democracy, 37–8; 'totalitarian democracy', 38–9
Dependence effect, the, 204–7; advertising, 205–7; consumer demand and behaviour, 205–7
Deprivation feelings and ethnic intolerance, 162–3
De Ruggiero, 39
De Seversky, A. P.: and Heartland theory, 411–12; on air power, 411
Dewey, John, 245
Dharma, 272, 509, 510, 512
Dickinson, R. L., 128
Dictatorship, 453–5: and democracy, 420–4, 446; dictatorial leadership, 449; perpetuation of system, 452
Diderot, 25
Di Lando, Michele, 526
Discourse on the Origin of Inequality (Rousseau), 25
Disraeli, 278
Doozoku pattern (Japan), 178
Duesenberry, Professor, 205
Duncan, J. Mathews, 124
Durkheim, Émile, 563, 564, 683
Duverger, study of political parties, 460

Early sex training, 114–17
East Pakistan refugees in Calcutta, 670–71
Easter Island, 134
Eastern Bengal, high natural population increase, 622, 626–8, 635, 636, 640
Eastern Europe, 395: Euro-Asia, 398–402; political divisions, 396, 397
Economies of Location, The (Lösch), 375
Egalitarianism, 37–8
Ego defence, 168–72
Egypt, population pressure in, 646, 647, 653, 662
Einaudi, Luigi, 34
Elizabeth I, 502
Elmira study of 1948 election campaign, 674, 677, 679–84
Emerging self, psychology of the, 15–22: methods of evaluating knowledge, 16–18; presentation of goal-directed behaviour, 18–22; study of human behaviour, 15–16
Émile (Rousseau), 39
End Social Inequality (Wootton), 280, 284
English Journey (Priestley), 278
'Eobiont' ('dawn life'), 6
Essays on Unsettled Questions (J. S. Mill), 259
Ethnic tolerance, 160–7: Negro attitudes, 161–3, 165–7; attitude towards controlling institutions, 164–6; controls, 163–6; distribution of intolerance, 161; external control, 164–5; feeling of deprivation, 162–3; religious attitudes, 164; social mobility, 161–3; stereotyped thinking, 166–7; superego control, 164
Euro-Asia, conception of, 398–402: continuous land-mass, 398, 400; favourable to mobility of horsemen and camelmen, 400–1, 407; Heartland, *see* separate entry; marginal regions, 400, 409; pivot area, 402–3, 407–8; Rimland theory, 409, 412–13; river drainage, 399; seapower and land-power, 400–2
European Economic Community, boundaries and frontiers within, 416, 417–23: Alpes-Maritimes, 419–21; Alsace-Lorraine, 418; areas of poor resources, 418; Franco-Italian boundary, 417–18; Limburg appendix, 418, 421–2; Low Countries boundaries, 418; northern France–Flanders boundary, 418; partitioned natural resources, 418–19
Evaluation of knowledge, methods of, 16–18: authority, 16–17; common sense, 16; empirical evidence, 16–17

Evolutionary history of man, 3–15: basic molecules of life, 4–5; evolution in action, 9–10; evolution of man, 13–14; growing and dying living units, 7; increase in complexity, 11–12; increase in efficiency, 12; leap to modern man, 14–15; medium- and short-term changes in animal life, 9–10; rate of evolutionary change, 12–13; survival of the fit, 7–8; views on origin of life, 3

F scale, 172–4
Fable of the Bees (Mandeville), 26
Fahlbeck, Pontus, 451
Family-planning communication, 684–93: approach, 686–7, 692–3; bureaucratic timidity, 689; contraceptive methods, 685–6; cost, 692; high-parity, 688; implementation, 688–9; mass-media, 689–90; nature of society, 685–6; place of research, 692; short-run programmes and research, 687–8; social-economic status, 688
Family size in underdeveloped areas, 655–67: contraception, 688–9; familial and kinship units, 660–1; industrialisation and urbanisation, 662–3; intermediate variables, 657–8, 667; literacy, 663–4; mass education, 663–4; mass media, 664; mortality rates, 659–60, 662; range of fertility rates, 665–6; sample surveys, 655–6, 665–6; social and economic changes, 662–3
Feeding situation: apathy and apprehension, 109–10; colic and recurring hunger, 111; conflicts and attitudes, 109–11; fear of being alone, 110; lack of social feeling, 110; secret learning of early years, 111; sociability and 'love', 110; unlabelled early conflicts, 111; weaning, 110–11
Feral children, 143–53
Ferri, Enrico, 527, 532, 534
Fertility history of U.S.A., 611–16: changes in family size, 613; changes in marriage patterns, 613–4; comparison with Canada, 615; hypotheses about increase, 614–5; total rate and productivity, 612–13
Fiji, population density, 646
Firmin, Thomas, 504
Flat land, 298–9
Fletcher, Ronald, 200
Flugel, J. C., 282
Foerster, V., 12
Ford, E. B., 10
Four Freedoms, the, 34
France: Church and State struggle, 533; interest groups, 467–8; political system, 465–6
Franco-Italian Convention (1951), frontier zone provisions, 417, 419, 421
Franklin, Benjamin, 503, 643
Frenkel-Brunswik, 172
Freud: and cleanliness training, 112; and Oedipus situation, 116; and the super-ego, 114
Friedman, Milton, 261
Fromm, Erich, 523
Frontier zones and boundaries, 413–23
Fulani, the, and Fulani emirates, 183–5, 187, 191
Full employment, 250–5
Full Employment in a Free Society (Beveridge), 252
Fullarton, John, 260
Fusataro, Takano, 234

Galápagos Islands finches, 9–10
Galsworthy, John, 282
Gandhi, and Gandhians, 508–11, 625
Gases of the early atmosphere, 4
General systems theory in geography, 65
General Theory of Employment Interest and Money (Keynes), 248, 253, 254, 256, 257, 262
Genes, 8: adjustment of frequencies, 11; and characteristics of the blood, 10
Ghenghis Khan, 399
Genotype, 8, 12
Ghurye, Professor, on Hindu *jātis*, 271
Ginsberg, Morris, 276
Glass, David V., 281, 664
Goal-directed behaviour, presentation of, 18–22: adjustment mechanisms, 21; environment, 19–20; general goal of well-being, 21; intensity of approach, 21; need for love, belongingness and psychological safety, 20; orientation, 20; physiological goals, 20; relationships with others, 22; schematic presentation, 18–19, 22
Goebbels, 452
Göhre, Otto, 532
Gonads and gonadotrophins, 121–2
Göring, 452
Gosplan, 227
Goths, the, 397, 398
Graziadei, Antonio, 534
Grotius, 37
Group leaders, struggle among, 525–35: demagogy, 525–6; maintenance of discipline within groups, 531–2; old leaders and aspirants to power, 528–31; oligarchy issuing from democracy, 527–8; popular revolutions and destruction of leaders, 526
Group norms, 170
Group pressure and influence on judgements, 578–85: analysis of individual differences, 581–2; categories of reaction, 581–2; 'compromise' partners, 584; experiment and first results, 579–81; experimental variations, 582–5; independent and yielding subjects, 581–2; majority size, 584; non-unanimous majorities, 582–4; stimulus-situation, 585; 'true partners', 583–4
Guesde, Jules, and Guesdists, 531, 532
Guest, Robert, 514, 515

Haggett, Peter, on location of zinc smelting, 318, 319
Haldane and Oparin, view of origin of organic molecules, 3

Hall, A. R., criticism of Heartland concept, 411–12
Hanau, 527
Harijan, Gandhian weekly, 511
Hartshorne, Richard, 415
Hauser, Philip M., 672
Haushofer, Karl, 394, 413
Heartland concept and the Heartland debate, 409, 412–13: Anglo-America as another Heartland, 411–12; point of gravity, 412–13; Rimland theory, 409, 412–13
Hedin, Sven, 408
Hegel, 39: view of the State, 29–30
Herbertson, A. J., and natural region hypothesis, 414
Hervé and his followers, 527, 531
Hess, Rudolf, 452
Hexagonal lattices, 367–71
Hicks, Sir John R., 249, 250
Hidden Dimension, The (Hall), 489
Hill, Stycos and Back, 658
Himmler, 452
Hindus, 627, 628, 631, 633–5, 670, 671: communal parties, 508–11, 513; Mahasabha, 508, 510, 511
History of European Liberalism (De Ruggier), 39
Hitler, 432, 433, 452
Hobbes and the Hobbesians, 23–6, 30: and limits of State action, 32; demand for strong State, 31; identification of State with legality, 23; on liberty of the subject, 31; 'state of nature' argument, 24, 26, 31; view of 'liberty of the Commonwealth', 30
Hogarth, D. G., 407–8
Hogben, Lancelot, 284
Holdich, Sir Thomas, 405–6, 408
Homo sapiens, 14
Homosexuality taboo, 115
Honke (main family, Japan), 177
Hooson, D. J. M., on Heartland concept, 412–13
Hoover, E. M., 335
Hopi, the, 129, 132
Hopkins, Harry, 573
Hovland, Carl, 692
Hoyt, H.: on land values in Chicago, 341–2; sector model, 354–7
Huichol, the, 131–2
Human behaviour, study of, 15–16: motivation for learning, 16
Hume, David, theory of money, 258–9
Huns, the, 397
Hunter, Rev. Henry, 297

Identification, 155–6
Ie (Japanese family), 175–7, 180, 182, 183
Ifugao head-hunters of the Philippines, 124, 133
Ila-speaking peoples of Africa, 133–4
Individuals, development in, 121–37
India: caste system, 271–5; family size, 659–61; political protest and urbanisation, 669–72
Indian political behaviour, 507–13: Hindu communal parties, 508–11, 513; law and civil disobedience, 510–12; struggle for power, 510
Indian population problems, 645, 646, 648, 651–3: community development programmes, 640–41; family planning and birth control campaigns, 637–8; industrial activities, 641–2; malarial regions, 637–9, 645; morale in peace and war (1931–47), 633–5; social and psychological significance of variability, 620–36; types of variability, 621–31; urban increase, 640–41
Industrial Development Certificates, 332
Industrialisation and democracy, 522–4
'Inertia', theory of social order, 565
Inflationary equilibrium, 259
Inhibition of anger, 119–20
Initiation rites for boys, 127–8
Interest groups and the political process, 459–71: and bureaucracy, 469–70: and political parties, 469–70; general objectives, 460–1; legislative process, 467–9; public opinion, 463–4; single-party and multi-party systems, 460; specific groups, 462–3; Western and non-Western systems, 470–1
'Interest' theory of social order, 562–3
Internalisation, 156–7
International Monetary Fund (I.M.F.), 263
Internationals, Communist, 456–9: Third International, 457
Ireland, government and industrial location in, 332–5: concentrations of new employment, 334; Industrial Credit Company, 333; Shannon Free Airport Scheme, 333–4; special development areas, 335
Islam, *see* Muslims
Isoo, Abe, 234
Italian Peace Treaty (1947), frontier zone provisions, 417
Italy: Constitution and social rights, 34; Fascism, 454, 455; Liberalism, 34; political system, 465–6

Jahoda, Marie, 277
Jainism, 274, 625
Jamaica: family size, 666; population pressure, 651
Jamma Praja Parishad, 511
Jan Sangh Party, 508, 510, 511, 513
Japan: birth rate, 661, 662; decline in mortality rate, 662; industrial expansion, 653; labour–management relations, 229–35; reduction of agricultural population, 653
Japanese Factory, The (Abegglen), 230
Japanese family, the, 175–83: adaptation, 180–1; child training and socialisation, 182–3; Civil Code, 177; co-ordination, 181; deviance, 183; integration, 181; kinship terminology, 176; marriage, 178–80; revised code, 177; stem and branch families, 177–8; trustee family, 175
Jātis (Hindu sub-castes), 271, 272, 274

Jaurès, Jean, and his followers, 527, 531, 532
Java man, 14, 18
Jayadeva, author of 'Indian Song of Songs', 627
Jennings, Sir Ivor, 428, 429, 431, 441
Jihad, the, 184
Jivaro of Ecuador, the, 129
Jones, Professor Wood, on similarities between man and Prosimian, 13
Joseph and His Brethren (Mann), 17
Jukun, the, puberty ceremonies, 126
Junod, on Thonga puberty ceremonies, 125
Justice and the State, 27–9

K hierarchy principle: fixed-, 369–71, 375; variable, 375
Kalecki, Michal, 253
Kanner and Mahler, study of autistic children, 147–50, 152: 'Case 9', 148
Kant, Immanuel, 35, 36
Karapatri, Swami, 510
Karma, See *Khama*
Kazak, the, 132
Keraki, the: as restrictive society, 130; initiation rites, 126–7
Kette, Max, 534
Keynes, J. M. (Lord Keynes), 205: and full employment, 251–5; monetary theory and Keynesian economies, 246–50, 256–8, 260–3
Khama (Karma), 272, 631
Khazars, the, 397
Khrushchev, N. S., 224, 452
Kinsey reports, 114, 130, 131
Kirov, 452
Kiwai Papuans of New Guinea, 129–30
Klein, Viola, 201
Knodel and Tangoantiang, on fertility and illiteracy, 663
Koran, the, 663
Koruhauser, Arthur, 521, 522, 524
Koruhauser, William, 513
Kshatriyas (Hindu caste), 271, 273, 509, 510
Kulturkampf, the, 533
!Kung Bushmen, relief of social tensions among, 43–106: absence of stealing, 104; artifacts, 104–5; avoidance of gossip, 97; food and hunting, 94–101; forms of verbal expression, 95–6; gift-giving, 95, 101–4; good manners, 96–7; restraint of temper, 97; taboo subjects, 95; talking and 'talks', 94–6
Kurd, the, 130
Kwoma, the, 128

Labour–management relations in Japan, 229–35: development of labour movement, 233–5; engagement of new employees, 230; 'familism', 232; strike action, 233; terms of employment, 230; trade unions, 229–30, 233–5; wage structure, 231; welfare services, 231–2
Lack, David, 9
Land-power and sea-power, 400–2, 407
Land rent, 289–90
Land values: land-surface value concept, 341–3; maps, 343–4; variables, 344–8
Lane, Robert E., 522–4
Laski, Harold J., 283, 286–7, 606
Latimer, Bishop Hugh, 505
Laud, Archbishop William, 502, 503
Law and order, 23–6: depreciation of order, 24–6; invitation to anarchy, 25–6; justification of the State, 23–4; political realism and legal positivism, 26; value judgements, 26; value of order, 24
Law and the Constitution, The (Jennings), 429
Lawrence, Sir John, 634
Le Bon, school of crowd psychology, 450
Lenge, the, puberty ceremonies, 125–6
Lenin, 452, 454, 457, 458: fight for centralised revolutionary 'élite', 443
Lepcha of India, the, 133
Lesu, the, 133, 134
Leviathan (Hobbes), 31
Lewis, Austin, 530
Lewis, R., 276
Liberation, political theory of, 32–4
Liberty of the Ancients Compared with that of the Moderns, The (Benjamin Constant), 35
Limburg appendix frontier zone, 418, 421–2
Lincoln, Abraham, 36
Lineage system among Tiv, 48–6
Lingayats, the, 273, 274
Lipset, Seymour, 513, 514, 522–4
Literary conception of history, 395
Lloyd James, A., 281
Locke, John, 29, 32–4, 36: chief philosopher of negative liberty, 32–4; on law of nature and of reason, 32; on natural rights of man, 32
Lomasny, Martin, 569
Lord, Russell, 17
Lorenz, study of varieties of animal behaviour, 149
Lösch, August, 375, 414–17
Low Countries boundaries, 418
Lydall, H. F., study of changes in distribution of personal income, 264, 265, 268–70

McGraw-Hill Plant Site Survey, 337–9
Machiavelli, 25, 26, 38, 429, 510: demand for strong State, 31; view of political liberty, 30–1
MacIver and Page, on the modern family, 199, 200
MacMunn, Sir George, 634
Magna Carta, 33
Magyars, the, 397
Mahan Capt. (Admiral) A. T., 401, 409, 413
Making-out (a counter-system of workers' control of work situation and relationships), 235–45: intergroup co-operation and relations, 244–5; major job categories, 240–1; new rules and systems, 241–4; operator interaction with service groups, 239; piecework, 240; rework,

240; shop groups, 238–40; time study, 240
Malaria eradication and control, 637–9, 645
Malenkov, 452
Mallet du Pan, 37
Malthus and Malthusianism, 640, 643–7
Mandeville, Bernard, 26
Man–environment system and innovation (coffee-growing in Chimbu), 65–79: centrality, connectivity and the network, 76; evolution of central place, 75–8; innovation path, 66–75; pre-colonial system, 66; preconditions of major innovations, 66, 68
Manifest and latent functions, 567–574
Mann, Thomas, 17
Manus, the, 128
Mao Tse-tung, 457
Maori, the, 134
Maratha, the, 624, 625, 627
Marginal purchase and marginal utility, 203, 207, 252
Marketing or supply principle in central-place theory, 366–9
Markham, Sir Clements, 408
Marriage: among Wodaabe, 189–99; changes in patterns in Britain, 199–202; in Japan, 178–80; in the U.S.A., 613–4
Marshall, Alfred, 248, 252, 257
Marshall, T. H., 276
Marx, Karl, and Marxism, 25, 458–9, 522, 527, 534, 565: in India, 508, 669, 670; Marxist criticisms of Keynes's theory, 253
Masaniello, 526
Mashruwala, K. G., 511
Massachusetts, theocracy of, 503
Masturbation taboo, 114–15
Maturation, physical, 121–3: primary and secondary sex characters, 122–3; puberty, 122–3
Maude, A., 276
Mauritius, population pressure, 646–8, 651
Meinig, D. W., on Heartland and Rimland concepts, 412–13
Mendel, and inheritance of genes, 8
Menstruation, 123–6
Mesopithecus, 13
Mess, H. A., 276
Metabolic cycles, 7
Micaud, Charles, 443
Michigan Completion Test, 172–4
Micro-economic problems in Soviet economy, 223–8
Midnapore, wolf girls of, 143–9
Milk Marketing Board, 297
Mill, John Stuart, 34, 37: analysis of working of credit expansion, 260; defence of individual against pressure of social forces, 32; monetary theories, 259–63; principle of utility, 32
Miller, Alice Duerr, 278
Millerand, Alexandre, 532
Mills, C. Wright, 523
Milton, John, 502
Minghi, on normal boundary contacts, 415
Miocene apes, 13
Miraculous Birth of Language, The (Wilson), 282
Mirkine-Guetzévitch, 436
Mixed-class marriages, children of, 120–1
Mobility aspirations and conflicts, 120–1
Molecules: adsorption theory, 4; basic molecules of life, 4–5; complex organic pigment molecules, 5; macromolecules, 5, 7, 8; nucleic acids, 4–5, 7; origin of organic molecules, 3
Monetary theory and Keynesian economics, 246–50: classical theory, 246; emphasis on level of employment, 246; full version of Keynesian theory, 248–50; indeterminacy, 247; IS and LM curves, 249–50; 'Keynesian Cross', 248; price level and economic fluctuations reduced to minor position, 246–7; theory of income determination, 247–8
Mongol invaders, 397–400
Montagu, M. F. Ashley, 124
Montesquieu, 428, 434, 436, 460: fable of the Troglodytes, 24–6; on moderate governments and political liberty, 35
Mookerjee, Syama Prasad, 511
Movement minimisation and location of industry, 319–29: blast furnaces, 320–2, 327–8; coal, 326–8; coke plants, 325–6; loss of weight during manufacture, 321–3; milk manufacturers, 320; open-hearth steel works, 327–8; sugar-beet factories and arable land, 319–21, 325; weight of material per operative, 323–4, 326–7
Mukerjee, Sir Asutosh, 627
Murngin of Australia, the, 128
Muslims, 622, 627, 628, 631, 633–5, 640, 670, 671
Mussolini, 433
Mysore, study of mortality rate in, 659–60, 665–6

Nama Hottentots, the, 132
Narayan, Jayaprakash, 510
Natural region hypothesis, 414
Natural selection, 7–8, 11: genes and chromosomes, 8
Nature and convention, 27–30: contrast between convention and nature, 29; differing concepts of the State, 27–9; interpretations of the State's origins, 28–9; justice and the State, 27–9; modern theory of the ethical State, 29–30
Nazis, the, 455
Neanderthal men, 14, 18
Negative liberty, 30–5: conflict between Individualism and Socialism, 34; Declarations of Rights, 33; difference from positive liberty, 35; liberal conception of the State, 32–4; limits of State action, 32, 34; removal of obstacles, 31, 34; 'social rights', 33–4; sphere of individual independence, 31, 32
Nehru, Jawaharlal, 511

Nicomachean Ethics (Aristotle), 27
Niewe Tijd group, 534
Nomadic peoples, 397–402: mobility, 400, 408; reasons for their overflow, 405, 408
Notestein, Frank, 685
Nucleic acids, 4–5, 7, 8
Nyae Nyae region of South Africa, 93

Oaken Heart, The (Allingham), 280
Obedience and disobedience to authority, Yale experimental programme, 594–608
Oedipus complex, 115–17: 'resolution', 117
On Liberty (Mill), 32
One Hundred Years of Land Values (Hoyt), 341
Oregon timber industry, 335–6
Origin of life, views on, 3
Oró, and repeating bases in nucleic acid structures, 4
Out of the People (Priestley), 278
Overpopulation, 648, 651–2, 654
Owen, Robert, 504

Paine, Thomas, 434
Paish, F. W., study of changes in distribution of personal income, 264–7, 269, 270
Pakenham-Walsh, Bishop H., 144, 147
Pant, Pandit, 511
Papago, the, 129
Parasites, life of, 12
Pareto, 25, 451
Parliament, supremacy of, 428–9
Parsons, Talcott, 199, 563, 565
Pasteur, Louis, 5
Peak, H., 171
Peking man, 14
Pennsylvania, tolerance in, 503
People's Choice, the, 674–77, 679, 681, 683
Peppered Moth (*Biston betularia*), 10
Permissive societies, 132–4, 137
Peròn ists, the, and Argentine army, 455
Persia and the Mongols, 399
Personal income, distribution of, 264–70: burden of taxation, 267–8; definition of personal income, 264–5; growth of forms of personal income, 268–9; incomes after tax, 266–7; incomes before tax, 265–6; percentile incomes, 265–6, 270; reasons for check to greater equality, 268–9; wages and salaries in manufacturing industry, 269
Petty, Sir William, 644
Phenotype, 8
Photosynthetic and respiratory cycles, 5–6
Physical maturation, 121–3
Pirie, N. W., 5
Pituitary gland and hormones, 121–2
Plato, 434
Plymouth, study of post-war industrial development, 330
Political machines, 567–74: and business interests, 569–73; and 'deprived classes', 569; bosses, 568–70; channels of social mobility, 570–1; key functions, 568; structural context, 568; subgroups, 569–73; vice, crime and rackets, 571–2
Political Parties (Michels), 448
Political parties and interest or pressure groups, 442, 453, 464–7: Anglo-American systems, 464, 470–1; France and Italy, 465, 467–8; non-Western systems, 464–5, 470–1; Scandinavia and the Low Countries, 465–8
Political parties, comparative study of, 440–59: and armed forces, 455; Communist parties, 443–4; definition of political party, 440–1; democracy and dictatorship, 440–4, 446; ideologies, 456; in pluralistic society, 452–6; internationals of parties, 446–9; link between government and public opinions, 442; selection of leaders, 442, 452; sociology, 448–52; towards classification, 444–8
Political protest and urbanisation, 669–72
Political systems, 424–7, 491–501: Aristotle's classification, 424–6; attributes, 491–3; boundaries, 492; concrete systems and abstract classifications, 427; demands, 493–5; differentiation within system, 493; identification 492–3; inputs and outputs, 493–9; integration of system, 493; multiplicity of criteria, 424; support, 495–500; units, 492; Weber's classification, 425–7
Politics (Aristotle), 27
Polo, Marco, 399
Polymorphism, 10
Ponape, the: as permissive society, 133; puberty ceremonies, 126
Popular sovereignty, 35–6
Population and resources, 648–54: current output and current population, 650–1; development of employment off the land, 653; 'disguised unemployment', 652; overpopulation, 648; resource conservation, 648–50; underpopulation, 653–4
Population and subsistence, 643–5
Population change: India in an era of change, 637–42; social and psychological significance of variability, 620–26
Population pressures in underdeveloped countries, 645–8: high population densities, 646, 647; rate of growth of population, 645–7; size of population base, 646, 647
Population trends in developing countries, 616
Positive liberty, 35–40: and popular sovereignty, 35–9; difference from negative, 35; threats from egalitarian democracy, 37–8
Potamic stage of civilisation, 401
Poujadism, 468
Poverty and Population (Titmuss), 280
P.O.W. camp, economic organisation of, 42–9; bully marks, 467; cigarette currency, 45; development and organisation of the market, 43–5; paper currency, 46–7; price-fixing, 47–8; price movements, 46; public opinion, 48–9
Praja Socialist Party, 508, 510, 513
Pressure groups, 442, 453. *See also* Interest groups

Priestley, J. B., 278–9
Primate evolution, 13–15
Princeton Office of Population Research, 663
Principles of Political Economy (J. S. Mill), 259, 260
Probiotic soup, theory of, 5, 6
Procunsul ape, 13
Prosimian (oriental tarsier), 13
Puberty, 122–7: and adolescence, 123–4; ceremonies in different societies, 124–7; precocious, 123
Puerto Rico, and family planning, 658, 659, 665
Pukapukas of Polynesia, 132
Punjab, the, 'colonisation' in, 622, 627–8, 633, 634, 636, 640
Puritanism, 501–7: collision with Crown, 502; contribution to political freedom and social progress, 504; economic interests, 502–4; ethical interests, 503–4, 506; moral self-sufficiency, 501; standards of social conduct, 502, 503
Pusztas, the, 397, 399
Pygmalion (Shaw), 281, 282

Quantity Theory of Money, 253–5, 258

Radek, Karl, 457
Railway freight rates in U.S.A., 293–6: class rates, 295, commodity rates, 295; freight-rate territories, 293–5; in-transit rates, 295–6; rate-group principle, 295
Rainborough, Colonel, 36
Ram Rajya Prashad, 510, 511
Rashtriya Swayamsevak Sangh (R.S.S.), 508, 510
Rational economic decisions, making of, 208–12: choice and maximisation, 208; constraints, 208–10; examples, 209–11; modern scientific management techniques, 208; objective function, 208; private and public interests, 211–12; social pay-off, 208
Ray, Ram Mohan, 627
Recreation, policies of, 535–59
Reference groups and attitude development, 585–94
Religion and the rise of capitalism, 501–7
Resource allocation and the pricing system, 218–19
Resource conservation, 648–50
Restrictive societies, 128–30
Reuther, Walter, 517
Rhode Island, tolerance in, 503
Ricardo, David, 207, 256–61, 263: Ingot Plan, 259; monetary writings, 257–61
Richards and Reining, study of the Bahaya, 656
Rienzi, 526
Rigola, Rinaldo, 535
Rimland theory, 409, 412–13: continental and maritime rimlands, 412
Robertson, Dennis, 262
Robespierre, 526
Robinson, Richard, 432
Robinson, W. S., 672
Roosevelt, Franklin D., 476, 573; and the Four Freedoms, 34
Ross and Hartmann, international comparison of labour disputes, 233
Rossi, Pellegrino, 430
Rousseau, Jean-Jacques, 36–9, 430: and the general will, 38–9; and the happy savage, 25; presentation of democracy, 37; seeds of totalitarianism, 38–9
Rovere study on interpersonal influence and communications behaviour, 674, 676–7, 680–84
Rowntree and Carrier, study of stability of marriage, 201
Roy, M. N., 508
Rueff, Jacques, 261
Rural land use, 289–315: agricultural production, 290–3; behavioural factors among farmers, 302–12; cash-grain farming and landforms, 297–301; geographical change in modified zonal system, 314–15; isolated state concept, 289–90; railway freight rates, 293–6; spatial variation in costs and revenues, 313–14
Rural Settlement and Land Use (Chisholm), 297, 315
Russell, Bertrand, 17, 18
Russia: and the Mongols, 399; central strategical position, 402, 407–8; eastward swoop across Asia, 401; Euro-Asia, 398–402; forest region, 396–8, 401; mobility of land-power, 402, 407; physical contrasts, 395–6; railway system, 402, 407; steppes, 396–401

Sanpoil, the, 130
Sanskritisation, 274, 275
Saracens, the, 400–1, 406: navigation of oceanic coasts, 401
Satiable wants, law of, 203
Satyagraha, 510–12
Schattschneider, E. E., 444
Schippel, Max, 532
Schlebecker, J. T., 315
Scott, J. P., 599, 606–7
Sea-power and land-power, 400–2, 407: discovery of Cape route to Indies, 401; Suez Canal, 401
Second Treatise of Government (Locke), 32
Secret of Hegel, The (Stirling), 428
Segmentary social structure, 51
Seljuk Turks, 400
Semi-restrictive societies, 131–2
Sen, Katayama, 234
Seniang, the, 132
Service centres and the distance factor, 384–92: criteria of centrality, 388–9; evolution of service centres, 390–1; population and centrality, 390; population distribution, 385–8; spatial pattern of hierarchy, 389–90; study of Southwestern Wisconsin and Southern England, 385–92
Sex training, early, conflicts produced by, 114–17: heterosexual approach responses and fear, 115–16; homosexuality

taboo, 114–15; masturbation taboo, 114–15; sex typing of personality, 115; specific genital anxiety, 116–17; unlabelled heterosexual conflict, 117
Sexual behaviour in human beings, development of, 127–34: American society, 130–1; permissive societies, 132–4, 137; reflexive components, 127–8; restrictive societies, 128–30, 137; semi-restrictive societies, 131–2
Sexual behaviour in subhuman species, 134–5, 137: lower mammals, 134-5; subhuman primates, 134
Sexual reproduction, 8
Shakespeare, 502
Shannon Free Airport Scheme, 333–4
Shaw, George Bernard, 281–2
Shudras (Hindu caste), 271–2
Shūsui, Kōtoku, 234
Sibling rivalry, 119
Sickle-cell anaemia, 10–11
Silent Language, The (Hall), 489
Siriono, the, 129, 132
Smith, Adam, 256, 258, 289
Snohomish County, Washington, study of hierarchy of central places, 376–83
Social conflict, functions of, 574–6
Social Contract (Rousseau), 25, 37, 39
Social influence, three processes of, 154–60: antecedents and consequents, 157–60; compliance, 155; identification, 155–6; internalisation, 156–7
Social mobility and ethnic tolerance, 161–3
Social order, problem of, 559–66: aspects, 559–60; coercion theory, 560–2; consistency, 559; 'inertia' theory, 565; 'interest' theory, 562–3; meanings, 559–60; persistence, 559–60; predictability, 559–60; reciprocity or mutuality, 559; restraint, 559; value-consensus theory, 563–5
'Social rights' concept, 33–4
Social stratification in England, psychological aspects of, 276–88: buyer–seller relationship, 278; choice of phrases, 281; clothes, 282; code of manners, 277; eating and drinking, 282; economic status, 277; education, 283; ideal of gentlemanliness, 283–4; 'North–South' difference, 281; opportunity-class, 280–1; religion, 283; snob-class, 280–1; speech, 281–2; wartime evacuation, 285–6; 'Working Class' and 'Ruling Class', 284–5
Socialist parties and trade unions, 465–7
Society of Friends, 504
Sociological theory, central problems of, 559–66
Sociology of political parties, 448–52: apprenticeship of democratic aspirants, 451–2; bureaucratisation, 449, 454; charismatic leadership, 449; concept of the masses, 450; degree of participation, 450–1; dictatorial leadership, 449, 452; dispensable occupations, 451–2; fluctuation in leadership, 451; functions, 450; institutional and personal leadership, 448–9; list of determinants, 449–52; selection of leadership, 452; size of organisation, 450
Sonia Shankman Orthogenic School, University of Chicago, 143, 147, 150, 152, 153
Sophists, the, 28
Soviet Russia: industrial expansion, 653; mammoth totalitarian apparatus, 454–5; micro-economic problems, 223–8; reduction of agricultural population, 653. *See also* Russia
Spinoza, 25
Sprott, W. J. H., 276, 277
Spykman, N. J.: criticism of Heartland concept, 409, 412, 413; Rimland theory, 409, 412
Stalin, 227, 432, 452, 454, 458
State, the: and justice, 27–8; and negative liberty, 30–5; 'conventionality' concept, 28–9; identified with legality, 23; interpretations of origins, 28–9; liberal conception, 32–5; modern theory of the ethical State, 29–30; 'naturalness' concept, 27–8
'State of nature' argument, 24, 26, 31
Steffens, Lincoln, 569, 570, 573
Steppes, the, and steppe people, 396–401, 406–7
Steytler on Chewa initiation rites, 127
Stoics, the, 28, 29
Strafford, Earl of, on Puritans, 502
Street corner society, 80–93: Cornerville and its members, 80–1; gang members, 81–4; gangs and girls, 84–9; individuals and the gang, 89–93; insignificant role of home, 89; Italian immigrants, 81; leader as focal point, 91–3; mutual obligations, 90–1; social activities away from the corner, 90; stable composition of group, 89, 90; use of money, 90–1
Subsistence and population, 643–5
Supplément au Voyage de Bougainville (Diderot), 25
Support in political system, 495–500: domain, 496–7; government, 496–7; mechanisms, 498–500; outputs, 498–9; political community, 496; politicisation, 499–500; quantity and scope, 497–8; regime, 496
Survival of the fit, 7–8
Sweden, study of farmers in, 302–12
Systems analysis, 208

Tagore, Sir Rabindranath, 627, 631, 634
Taiwan, family planning in, 659, 661–4, 666, 667, 690, 691
Tamerlane, 399
Tartar nomads, 401
Terman, study of premarital intercourse in U.S.A., 131
Thaer, Albrecht, 289
Thalassie stage of civilisation, 401
Third International, 457
Thonga, the, puberty ceremonies, 125
Thornton, Henry, analysis of working of credit system, 260–1

Thouless, R. H., 276–8
Threshold population, 363–5, 372
Tikopia, 132
Titmuss, Richard M., 264, 280
Tiv, the, social and political organisation of, 472–77: age sets (*Kwagh*), 476; law, 475–6; lineage system, 472–7
Tocqueville on egalitarianism, 37, 38
Tooke, Thomas, 260
Totalitarian democracy', 38–9, 442–3
Toynbee, Arnold, 458
Traffic principles in central-place theory, 367–70
Training situations, critical, 109–21: anger–anxiety conflicts, 118–20; cleanliness training, 111–14; early sex training, 114–17; feeding situation, 109–11; frustrated mobility aspirations and aggression, 120–1
Treatise on Money (Keynes), 257, 262
Treaty of London (1839), 418
Treaty of Paris (1815), 418
Treaty of Rome and internal frontier zones, 417, 419, 423
Treaty of Turin (1860), 417
Treaty of Utrecht (1713), 417
Treaty of Vienna (1815), 418
Tree-shrew, oriental (*Tupaia*), 13
Trobrianders, the, 124, 132, 133
Trotsky, 452
Truk, 128, 133: initiation rites, 126
Turanian nomadic invaders, 397, 401
Turati, Filippo, 527
Turks, the, 399, 407: Seljuk Turk control of Nearer East, 400

Über den Standort der Industrien (Weber), 316
Ulbricht, Walter, 457
Underdeveloped societies: and population pressure, 645–8; norms for family size, 655–67
Underpopulation, 653–4
United Auto Workers (U.A.W.), 517
Untouchables, 271–3
Urban land use, 340: models, 349–61
U.S.A.: as restrictive society, 130–1; caucus system in Congress, 531; federal system, 467; fertility history, 611–16; railway freight rates, 293–6; separation of powers, 467; social stratification, 276
U.S.S.R., *see* Soviet Russia
Utility: and demand, 203–4; law of diminishing utility, 203, marginal, 203, 207, 252; total, 203

Vaillant, 527
Vaishyas (Hindu caste), 271–2
Value-consensus theory of social order, 563–5
Varangians, the, 398
Vasco da Gama, 394
Vedda, the, 130
Venda, the, puberty ceremonies, 126
Vikings, the, 398, 406: navigation of oceanic coasts, 401
Viviani, René, 532
Von Vollmar, 527

Wagman, M., 172
Walker, Charles, 514, 515
Wallace, A. R., and survival of the fittest, 7
Wapisiana, the, 129
Ward, Barbara, 669
Weaning, 110–11
Weber, Alfred, least-cost industrial location model, 316–22, 328–30, 335–8
Weber, Max: and dispensable occupations, 451; and rational legal authority, 512; classification of political systems, 425–7; on stages of party development, 444
Weimar Republic, the, 465
'Welfare optimum' and the pricing system, 216–17
West Indies, population pressure in, 646, 647
Western Bengal, stagnation of population, 622, 625–6, 628, 634–6, 639–41
Wheare, Sir Kenneth, 429, 431
White, William Allen, 453
Whitman, Walt, 279
Wilkinson, Spencer, 401, 404–5, 409
Wilson, Bryan, 199–200
Wilson, Richard Albert, 282
Wilson, Woodrow, 568
Wingo, L., 340
Wodaabe of West Bornu, the, 183–202: bride-removal and 'sleeping out', 190–2; first pregnancy, 192; formation of family and herd by first marriage, 189–94; 'home-building', 193–4; homestead and family, 185–9; interdependence of family and herd, 184–5; male and female spheres of action, 187–9; marital contract, 194–5; naming of first child, 192–3; preliminaries and betrothal, 189–90; values of marriage by betrothal, 196–9
Wogeo, the, 133
Wolf boys: of Agra, 138–43, 149–50; of Aveyron, 143
Wolf girls of Midnapore, 143–9
Wootton, Barbara, 276, 279–80, 284
Workers' control of work situation and relationships, 235–45

Yale University experimental programme on obedience and disobedience to authority, 594–608: acquired behaviour dispositions, 599; background authority, 602–3; closeness of authority, 599–600; denial and narrowing of cognitive fields, 598; empathic cues, 597–8; general laboratory procedure, 595–6; immediacy of victim, 597–9; incipient group formation, 598–9; levels of obedience and defiance, 604–6; meaning of obedience and disobedience, 594–5; phenomenal unity of act, 598; pilot studies, 596; Proximity and Touch-Proximity, 597–9, 606; reciprocal fields, 598; Remote condition, 597–8; subject population, 595; tensions, 600–2; Voice Feedback, 597–8

Yermak the Cossack, 394
Yudkin and Holme, study of working wives, 201
Yugoslav iron and steel industry, location factors in, 330–2: Five-Year Plan (1947–51), 330–1; Cominform Revolution, 331–2
Yungar, the, 130
Zhadanov, 452
Zibordi, Giovanni, 534
Zimmerman: concept of trustee family, 175; on instability of modern family, 200
Zinc smelting, location of, 318
Zoning in cities, 340–61: concentric zone model, 349–53, 358; dynamic quality of land use, 358–61; land-use models, 349–61; land-value maps, 343–4; land-value surface, 341–3; multiple regression analysis, 344–8; nuclei model, 357–8; sector model, 353–8; urban land use, 340; variables, 344–8
Zuckerman, Sir Solly, 134
Zweig, F., 277